

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/147276>

**Copyright and reuse:**

This thesis is made available online and is protected by original copyright.

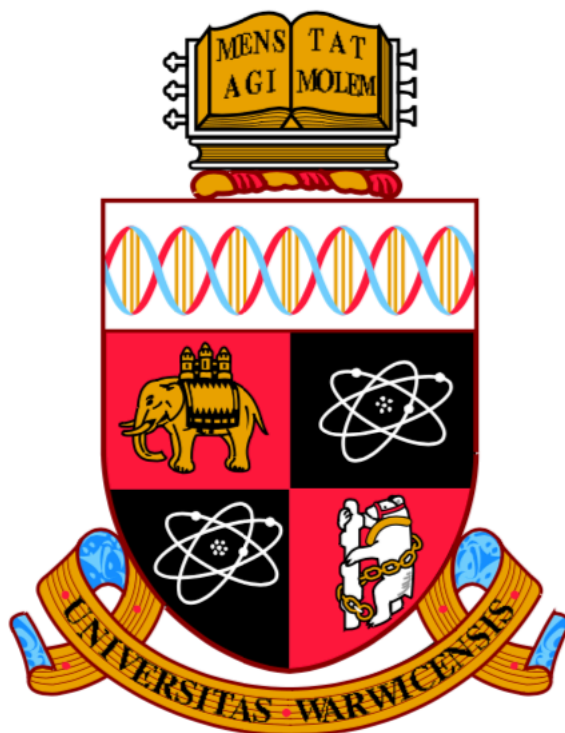
Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

# Development of Phage-assisted Evolution and Riboregulation Strategies



Eduardo Goicoechea Serrano

Thesis submitted to the University of Warwick  
for the Degree of Doctor of Philosophy

School of Life Sciences

---

October 2019

THE UNIVERSITY OF  
**WARWICK**

## Table of Contents

ACKNOWLEDGEMENTS.....	5
ABSTRACT .....	7
CHAPTER 1. INTRODUCTION.....	9
1.1 RNA. More than a middleman .....	10
1.2 Riboswitches, the smart way of regulating.....	12
1.3 Riboswitch improvement: Build-a-switch .....	14
1.4 Directed evolution: Playing God on a smaller scale. ....	16
1.5 Bacteriophages: Nature's most efficient killing machines .....	17
1.6 The T7 Phage: You've heard of him .....	18
1.7 Phage-assisted strategies.....	18
1.8 My project.....	21
AIMS AND MAIN OBJECTIVES .....	26
CHAPTER 2. MATERIALS AND METHODS.....	27
2.1 Original cell strains .....	27
2.2 Original phage strains.....	28
2.3 Lysogeny broth media .....	28
2.4 Theophylline solution.....	29
2.5 Saline-Magnesium (SM) buffer .....	29
2.6 M9 media .....	29
2.7 Antibiotics .....	29
2.8 Competent cell production .....	30
2.9 Transformations.....	31
2.10 Plasmids .....	31
2.11 Primers and Gblocks.....	35
2.12 Restriction enzymes.....	38
2.13 Ligation.....	39
2.14 Polymerase Chain Reaction.....	39
2.15 Goldengate assembly.....	39
2.16 Agarose gel electrophoresis.....	40
2.17 DNA fragment purifications.....	40
2.18 Bacterial culture .....	40
2.19 Phage killing/infection .....	40
2.20 Lysate filtration.....	41

2.21 Bacterial Growth assays .....	41
2.23 One-Step assay .....	41
2.24 Next generation sequencing.....	42
<b>CHAPTER 3. RIBOSWITCH LIBRARY ASSEMBLY .....</b>	<b>43</b>
3.1 Introduction.....	43
3.2 Specific Objectives.....	45
3.3 Specific materials and methods .....	46
3.4 Results: .....	53
3.5 Discussion of results: .....	57
<b>CHAPTER 4. RECOMBINATION-BASED EVOLUTION STRATEGY.....</b>	<b>58</b>
4.1 Introduction:.....	58
4.2 Specific Objectives.....	64
4.3 Specific materials and methods:.....	64
4.4 Results: .....	77
4.5 Discussion of results:.....	103
<b>CHAPTER 5. TRANSDUCTION-BASED EVOLUTION STRATEGY. ....</b>	<b>112</b>
5.1 Introduction.....	112
5.2 Specific objectives.....	116
5.3 Specific materials and methods .....	116
5.4 Results .....	122
5.5 Discussion of results.....	128
<b>CHAPTER 6. Q<math>\beta</math> PHAGE INDUCIBLE RNA SYSTEM .....</b>	<b>130</b>
6.1 Introduction: .....	130
6.2 Specific Objectives:.....	132
6.3 Specific materials and methods .....	132
6.4 Experimental procedure and Results: .....	135
6.5 Discussion of results .....	143
<b>CHAPTER 7. OVERALL CONCLUSION &amp; FUTURE DEVELOPMENTS .....</b>	<b>146</b>
7.1 Development of a phage riboswitch library .....	146
7.2 Development of strains for positive and negative selection of riboswitches.....	147
7.3 Development of an evolutionary system .....	148
7.4 Direct evolution of a riboswitch.....	148
7.5 Secondary objectives:.....	149
7.6 Closing Remarks:.....	150
<b>8 SUPPLEMENTAL MATERIALS: .....</b>	<b>152</b>
<b>9 BIBLIOGRAPHY.....</b>	<b>162</b>



Figure 1. Riboswitch domains and their function .....	12
Figure 2. Mechanism of action of several engineered riboswitches.....	14
Figure 3. Overview of PACE process.....	19
Figure 4. Overview of transduction process to extend phage host range. ....	20
Figure 5. Homologous recombination of the sequence of interest into T7 phages.....	23
Figure 6. Parental plasmid EG_001.....	33
Figure 7. Parental plasmid pSEVA631.....	33
Figure 8. Positive selection parental plasmid EG_002+.....	34
Figure 9. Negative selection parental plasmid EG_002-.....	34
Figure 10. Overview of One-Step plaque assay .....	42
Figure 11. Initial amplification for HR fragment. ....	47
Figure 12. Fixed and random riboswitch fragments inserted in HR fragment.....	48
Figure 13. Structure of the different packaging signals. ....	51
Figure 14. Homologous recombination plasmid carrying riboswitch variants. ....	53
Figure 15. Final Homologous Recombination fragment. ....	54
Figure 16. Results from HR procedure at different MOIs. ....	55
Figure 17. Sequence variation within obtained plasmid and phage populations.....	56
Figure 18. Genome and virion structure of T7 phage. ....	58
Figure 19. Schematic representation of the positive selection process. ....	61
Figure 20. Schematic representation of the negative selection process. ....	62
Figure 21. Diagram of the interactions between MMF, mmfR and mmyB. ....	63
Figure 22. Riboswitch versions. ....	65
Figure 23. Schematic representation of the original bioreactor. ....	69
Figure 24. General structure of the full phage evolutionary procedure. ....	70
Figure 25. Amplification steps for NGS preparation. ....	73
Figure 26. Results from cell killing assays in positive selection strains. ....	80
Figure 27. Results from One-step assays in positive selection strains.....	82
Figure 28. Result from cell killing assays in the negative selection strains.....	84
Figure 29. Specific growth rates of negative selection strains. ....	86
Figure 30. One-step assay for the negative selection step.....	87
Figure 31. qPCR amplification for Next Generation Sequencing. ....	89
Figure 32. Agarose gel of Indexing PCR. ....	90
Figure 33. Riboswitch sequence evolution throughout generations. ....	92
Figure 34. Virulence index example data. ....	97
Figure 35. Virulence index comparison between populations.....	98
Figure 36. GFP activation in the mmfR-mmyB/MARE-4 system. ....	101
Figure 37. T7 infection of mmyB/MARE4 strains.....	102
Figure 38. General structure of packaging signals. ....	116
Figure 39. General structure of the Phagemid evolutionary procedure.....	121
Figure 40. Number of colonies after transduction assays.....	125
Figure 41. Number of PFU after plaque assays with alternative lysates.....	126
Figure 42. Number of CFU after transduction assays with alternative lysates. ....	127
Figure 43. Q $\beta$ phage-based strategy overview.....	131
Figure 44. Diagnostic PCR for Q $\beta$ presence.....	135
Figure 45. Induction results. ....	137
Figure 46. Colony growth and diagnostic PCRs for Q $\beta$ presence. ....	138
Figure 47. Comparative growth of generations of Q $\beta$ -expressing cells.....	140
Figure 48. Q $\beta$ -GFP plasmid expression. ....	142

Table 1. List of <i>E. coli</i> strains used throughout the PhD.....	27
Table 2. List of original T7 strains used throughout the PhD .....	28
Table 3. List of antibiotics used throughout the PhD.....	29
Table 4. List of parental plasmids used throughout the PhD. ....	32
Table 5. List of primers used throughout the thesis. ....	35
Table 6. List of <i>E. coli</i> genes identified as essential for T7 growth. ....	60
Table 7. NGS sample nomenclature. ....	94
Supplementary figure 1. Intermediate recombinant plamid constructs.....	152
Supplementary figure 2. Positive control plasmid for positive selection .....	153
Supplementary figure 3. Negative control plasmid for positive selection. ....	153
Supplementary figure 4. Test plasmid structure for positive selection. ....	153
Supplementary figure 5. Negative selection plasmids .....	154
Supplementary figure 6. Homologous recombination plasmids for <i>mmfR</i> variants. ....	155
Supplementary figure 7. Selection plasmids for the <i>mmfR</i> strategy.....	156
Supplementary figure 8. Phagemid plasmid versions.....	157
Supplementary figure 9. Original Q $\beta$ IDT plasmid.....	158
Supplementary figure 10. Q $\beta$ -GFP plasmid. ....	158
Supplementary figure 11 . 3'UTR gene fragment for PURE experiments. ....	159
Supplementary Table 1 . List of T7 genes.....	159
Supplementary Table 2. Primer combinations for each NGS sample. ....	160
Supplementary Table 3. qPCR plate distribution of samples. ....	161

### ACKNOWLEDGEMENTS

I would like to thank all my family and friends for their continuous support throughout these 4 years. Whether at home or away, even if you are not here, I love you.

- My parents Eduardo, Eva, Javier and Ana for their never-ending encouragement, always believing in me and checking in on me, even when I was not in the best place
- My siblings, Javier, Marina, Lucia, Ignacio, Elvira, Paula, Guillermo and Eva, thank you for all the jokes and chats that made me feel that I was still close
- A Felipe, por ser el mejor abuelo
- To my Uncle Luis and my Aunt Isabelle, for their scientific advice and support
- To all my friends back home in AlcoSanse and beyond, Adrián, Alejandro, Alicia, Alfonso, Clara, Daniel, Eduardo, Elisa, Germán, Irene, thank you for putting up with my weirdness and “surprise” showing ups
- To the people from Los Chuchesh, for showing me that no matter how low one gets, there’s always a video of someone doing much worse
- To all my Madrileños, Angel, Belen, Carlos, Geles, Gloria, Alv-Kantala, Jesús, Juan (x2), Marta, Pako, Sergio and Sonia, for always having space for one more
- To the Fernández García family (And Mak) for being close despite time and distance
- To my Sifu & friends of the Xiang Xing Quan style, for always having space for one more
- To the denizens of 107 Greenwood Court, Andrea, Aurelija, Clare, Elena, Iulia, and Sami for being my 1<sup>st</sup> (and best) friends in Leamington and sticking up with me EVEN after that.
- Special thanks to Dr Aurelija Grygonite for being the best gym buddy/housemate/reviewer and putting up with THE DRAMA.
- To all the people from 118 Mason Avenue, Andrea, Anna, Dan, Rory, as well as Alonso, Cathal and Natassa, for being some of the best friends one could find.
- To Callus & friends, for showing me no one is too inept or far gone to do the right thing
- To Marina and Ruth, because never was a 24-h flight to Madrid more successful.
- To Rapo, for handling my jokes gracefully
- To the friends I see mostly electronically, but still help you smile, Andie, Alba, Crispy, Joak, Lunaticia, LuciRodher, Maria, MariLOL, Marina, Pedro, Sabri, Sami, Viki, ~~Xab~~
- A Patricha. Por que eres un solecito ☀.
- To the Warwick Magic Society, for some much-needed weekly breaks.
- To the Chronobiology lab, Helena, Kristin, Ewan, Laura, Lauren, Robert & Swati for making me feel like an honorary member
- To my friends from SLS. Steven for support, Tamsin for her help with this
- To Dr. Emma L. Denham, for her invaluable help and support
- To Shanan, for being a good friend even when forced to not see each other
- To Naomi. For standing on her own two feet and helping me do the same
- To Jessica, for solving mysteries and rewriting history
- To my labmates, Ariadna, Carlos, CleNira, Esther, Jesús, John, Manish, Maria (x2), Matt, Paul, Robert, Rui, Satya, and Tabitha, who went through this with laughs and counsel
- Special thanks to Paul McDonald, for coding, gaming and Fargo-ing.
- To all my colleagues from the MetaRNA consortium, for 3 years of scientific camaraderie
  
- To my supervisor, Prof. Alfonso Jaramillo, for giving me the opportunity to be part of this project, to learn and improve, and to jumpstart my scientific career
- To my advisory panel, Prof. John McCarthy and Prof. Munehiro Asally, for their help along these years.

*“When the going gets tough, the tough get going”* Joseph P. Kennedy

## DECLARATION

This thesis is submitted to the University of Warwick as part of my application for the degree of Doctor of Philosophy. It has been composed by me and has not been submitted in any previous application for any degree.

All work presented in this thesis (including data generation and data analysis) was carried out by the author (Eduardo Goicoechea Serrano) except where properly indicated and in the cases below:

- Figure 33 was produced by Carlos Bondía, from Manichahn lab, Vall d'Hebron Institut de Recerca, Barcelona
- Figure 36 and Figure 37 were produced by Dr Manish Kuswaha, from Micalis Institute, Paris
- Figure 48 was produced by Anis Senoussi, from André Estévez-Torres' lab, Sorbonne Institute, Paris

**ABSTRACT**

Riboswitches are short RNA sequences that undergo a structural modification in the presence of a specific molecule, regulating gene expression in the process. A few examples of riboswitches have been described in eukaryotes, but they are present in many species of bacteria. These molecules are of great interest in synthetic biology for their inductive regulatory properties, which could be used in metabolic studies, or as biosensors in medical, industrial or environmental contexts. However, riboswitches are very substrate-specific, creating a problem for their use with novel compounds; and current development procedures suffer from issues such as being too laborious and not using *in vivo* conditions. By developing strategies based on Phage Assisted Continuous Evolution (PACE) and using T7 phage along with a double selection system, we aim to develop a way of obtaining riboswitches that show improvements compared to a control sequence or with a different specificity. This procedure could even be used to develop novel riboswitches not present in nature. Once the method is properly established, the strategy could be used not only for the directed evolution of riboswitches, but also other types of sensors, such as protein or RNA receptors. Alternative evolution strategies based on phage transduction were also tested, along with inducible RNA systems based on the phage Q $\beta$ . These three processes have yielded different results, namely a randomised library of theophylline riboswitch sequences in phages, a working transduction system, and an inducible RNA plasmid, respectively. In the first case, by using a double selection process to achieve evolution, we have shown sequence variation within the phage populations progressively decreased over time, until a single sequence was prominently represented in the population. The sequence showed higher activation folds at a lower concentration of the activating molecule than in initial generations, taken as an indication that evolution had taken place. This library was also adapted into the transduction strategy thanks to the use of a working T7 phage packaging signal. As for the RNA-based plasmid system, we have shown it to be an inducible, transient expression system that could be used as a novel way to regulate gene expression and bypass CRISPR systems. This set of results speak of the possibilities held by these strategies, not just for their specific areas of research, but for synthetic biology at large, from bioremediation to treatment.

## List of abbreviations

PFU: Plaque Forming Units	CFU: Colony forming Units
Amp: Ampicillin	Kan: Kanamycin
Gent: Gentamycin	Cm: Chloramphenicol
MOI: Multiplicity of Infection	EOP: Efficiency of Plating
DNA: Deoxyribonucleic Acid	RNA: Ribonucleic Acid
<i>cmk</i> : CMP kinase	CMP: Cytidine monophosphate
<i>trx</i> A: Thioredoxin A	gp5: Gene product 5
DNApol: DNA polymerase	Rep: Replicase
RBS: Ribosome Binding Site	SD: Shine-Dalgarno sequence
MM/F: Methylenomycin/ furans	HR: Homologous Recombination
NGS: Next Generation Sequencing	qPCR: Quantitative PCR
PACE: Phage-Assisted Continuous Evolution	
PANCE: Phage-Assisted Non-Continuous Evolution	
LIMS: Laboratory Information Management System	
CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats	

## CHAPTER 1. INTRODUCTION

The ability to interact with the environment and react accordingly is one of the most basic abilities living organisms have (Malmstrom, 2010; Pearce, 2010). The biological elements allowing for this to take place are known as biosensors. From extremely complex photoreceptive systems allowing animals to see the world (Davson & Perkins, 2018), to the bacterial membrane receptor molecules reacting to the presence of nutrients or competitors in the environment (Braga, Dourado, & Araújo, 2016), these all share one basic trait: they take outside impulses and transform them into something the organism can identify and react to (Bhalla, Jolly, Formisano, & Estrela, 2016).

The ones that currently have the spotlight are, ironically, the least complex ones, working at the molecular level, based on conformational changes or molecule binding, who can regulate multiple processes, from metabolite production to gene expression, at the transcriptional or translational level. There is an exorbitant number of different biosensors in this category, as well as functions for them in the natural world (Mehrotra, 2016). They tend to be very specific, some are very conserved in evolution (Margoliash, 1963; Serganov & Nudler, 2013), while others are more specific to a smaller group of species.

The specificity, and in some cases simplicity, of these biosensor systems, have made them a very valuable target not just for research, but for uses such as industry, bioremediation, or medical treatments (Carpenter, Paulsen, & Williams, 2018). What's even more, thanks to synthetic biology, these natural biosensors have either been incorporated into organisms that did not originally produce them (Ravikumar, Baylon, Park, & Choi, 2017), had their efficiency improved via targeted mutations or evolution so that they can be used in alternative environments or processes (Bradley & Wang, 2015; Mahr et al., 2015; Stella, Wiechert, Noack, & Frunzke, 2019), or served as the base for the development of novel sensory circuits (B. Wang, Barahona, & Buck, 2013).

Possibly the most important development is the creation of artificial biosensors, using the natural ones as blueprints or examples (Rogers et al., 2015). These



studies have helped not only develop new molecules, or create sensors that detect compounds not naturally produced (Ramajani & Sayre, 2018), but have also shed light on the relationship between structure and function that lies at the core of some of these biosensors (Vallée-Bélisle & Plaxco, 2010); where a mutation can eliminate the function, modify its binding efficiency, or even create a specificity for other molecules.

Understanding the laws that govern these changes could lead to not just a better, deeper understanding of the subject, but also could harbour a completely new way of developing sensors with very specific objectives, which could greatly benefit research fields like infection treatments or industrial processes.

### **1.1 RNA. More than a middleman**

Nucleic acids are some of the basic elements that comprise living organisms. The most well-known one is obviously the double-stranded Deoxyribonucleic acid, or DNA. Although not as universally acknowledged as DNA, RNA plays an equally important role, being already associated with protein synthesis by 1939 (Caspersson & Schultz, 1939). Being a single-stranded molecule, and the middle ground between DNA and proteins, RNA was initially seen as a mere intermediary, but as research went on, more information was obtained on the actual level of regulatory control RNA holds in gene expression (Morris & Mattick, 2014). RNA is actually highly structured, and able to fold in on itself, thanks to self-complementary regions (Tinoco & Bustamante, 1999), which can create regions of double helix structures. These structural capabilities allow RNA to achieve multiple functions (Higgs, 2000), and showcase the fact that there is more than one type, with each serving a different, specific function. It is not within the scope of this research to delve too deeply into the categories of RNA, but to make special emphasis on mRNA.

Messenger RNA, or mRNA, was initially a theoretical concept within Francis Crick's "Central Dogma of Molecular Biology" (First stated 1958, properly published in Crick 1970), by which DNA is transcribed into mRNA, and this is then



translated into amino acids that form proteins. mRNA was isolated and proven to exist in 1958, thanks to research conducted by Hildegard Lamfrom and collaborators (Schweet, Lamfrom, & Allen, 1958); which started to unveil the level of intricacies within RNA biology, and all the roles its different types played in a functional organism. The most relevant one for the purpose of this research were its regulatory functions.

In recent times, RNA has, slowly but steadily, become a much-appreciated regulatory element to be used in transcription and translation of genetic sequences (Erdmann, Barciszewska, Hochberg, de Groot, & Barciszewski, 2001; Morris & Mattick, 2014). Regulatory RNA elements have more than enough qualities to stand their ground against protein elements given some of their positive properties:

- Lower metabolic cost and higher genomic economy, since regulation is at the genetic level.
- Highly conserved in evolution (Miranda-Ríos, Navarro, & Sobero, 2001)
- Specific recognition via base pairing.
- High specificity and orthogonality (Jenison, Gill, Pardi, & Polisky, 1994), which does not require high levels of the complementary molecule.
- Thanks to the above; modification of function is much faster and easier.

And despite some of their negative ones:

- Easy degradation and instability.
- Negative side of an easy modification of function/structure.
- Lower binding affinities than protein molecules.

All these have allowed researchers to adapt RNA molecules in order to form new regulatory circuits to be used *in vitro* or *in vivo* (Anzalone, Lin, Zairis, Rabadan, & Cornish, 2016), or development of new forms of therapy (Wagner et al., 2018)

## 1.2 Riboswitches, the smart way of regulating

One particular type of RNAs are known as riboregulators (Krishnamurthy et al., 2015) and comprise a group that includes aptamers (single strand oligonucleotides that change their structure in order to bind to proteins) (Hermann & Patel, 2000), or the RNA molecules known as riboswitches (Edwards & Batey, 2010; Serganov & Nudler, 2013). Riboswitches are regulatory mRNA molecules which can modify their structure due to environmental conditions, such as temperature changes, or the binding of a specific molecule in a segment of their sequence (Sherwood & Henkin, 2016; Tucker & Breaker, 2005). This changes the transcription or translation rate of sequences downstream from the riboswitch (Roth & Breaker, 2009) and is such a specific binding, that they can discern between specific molecules and closely related analogues.

Riboswitch have two distinct domains, an aptamer domain, which detects the specific ligand and produces a change in 3D conformation; and the expression platform, which affects gene expression based on such changes.

Figure 1 shows a diagram of these two elements and their function within a typical riboswitch.

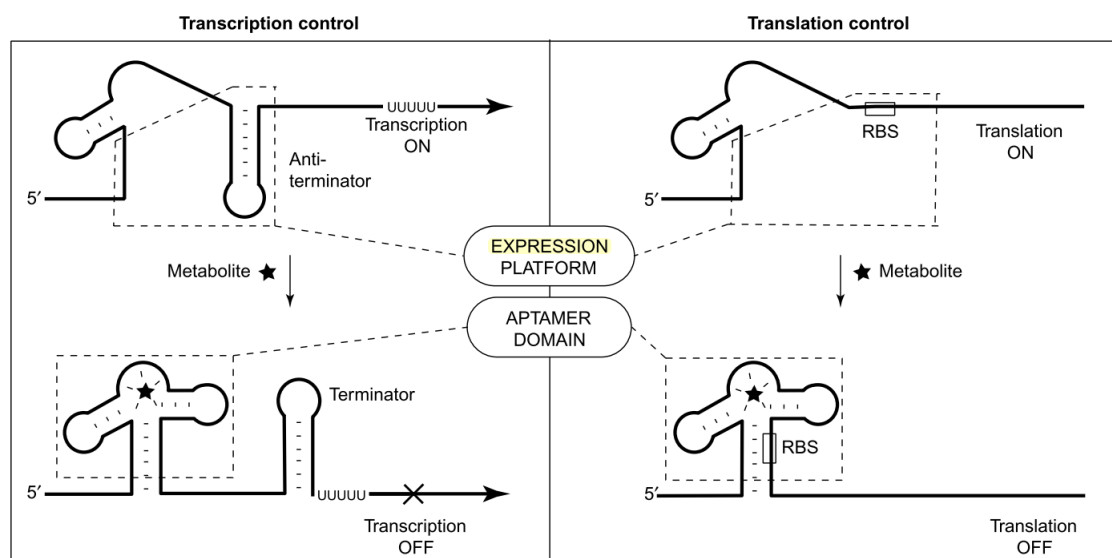


Figure 1. Riboswitch domains and their function.

(From Tucker and Breaker, 2005) Common mechanisms of riboswitch regulation, at the transcription and translational levels. As shown, in both cases, the structural change in

the aptamer domain after binding to a specific molecule affects the expression platform, and thus the outcome of the specific process.

Multiple organisms use riboswitches, mostly bacteria (Nudler & Mironov, 2004), but one specific riboswitch, known as the TPP riboswitch, has been detected in eukaryotes, such as plants or fungi (Sudarsan, Barrick, & Breaker, 2003). The fact that no riboswitches have been detected in mammals also raises the possibility of their use in human therapy, either as antibiotic targets (Blount & Breaker, 2006), or because research suggests they may not interfere with normal ligand interactions present in a mammalian cell environment, due to the very specific way in which the ligand binds to the switch (Montange & Batey, 2006). It is because of these reasons that riboswitches are not expected to cause any serious side effects in humans if used as therapy agents.

All these reasons have made riboswitches a very interesting group of molecules to work with, and given them a very promising lead in synthetic biology research and development (Machtel, Bąkowska-Żywicka, & Żywicki, 2016; Robinson, Medina-Stacey, Wu, Vincent, & Micklefield, 2016). A great deal of research has been done not just on the discovery (Bengert & Dandekar, 2004) and characterization of riboswitches, but also on their improvement and the development of synthetic riboswitches in both prokaryotes and eukaryotes (Findeiß, Etzel, Will, Mörl, & Stadler, 2017; Groher & Suess, 2014; Win & Smolke, 2008). This is due to riboswitches holding much potential for therapeutic (Mehdizadeh, Saeid, & Barzegar, 2016), metabolic engineering, or bioremediation (Pham et al., 2017) applications, and even for the more novel CRISPR technologies (Galizi & Jaramillo, 2019; Makarova et al., 2011). These can be obtained starting from a previously known riboswitch (Ceres, Trausch, & Batey, 2013) or aptamer that reacts against a specific molecule of interest. Figure 2 shows different ways in which artificial riboswitches can affect gene expression in prokaryotes and eukaryotes.

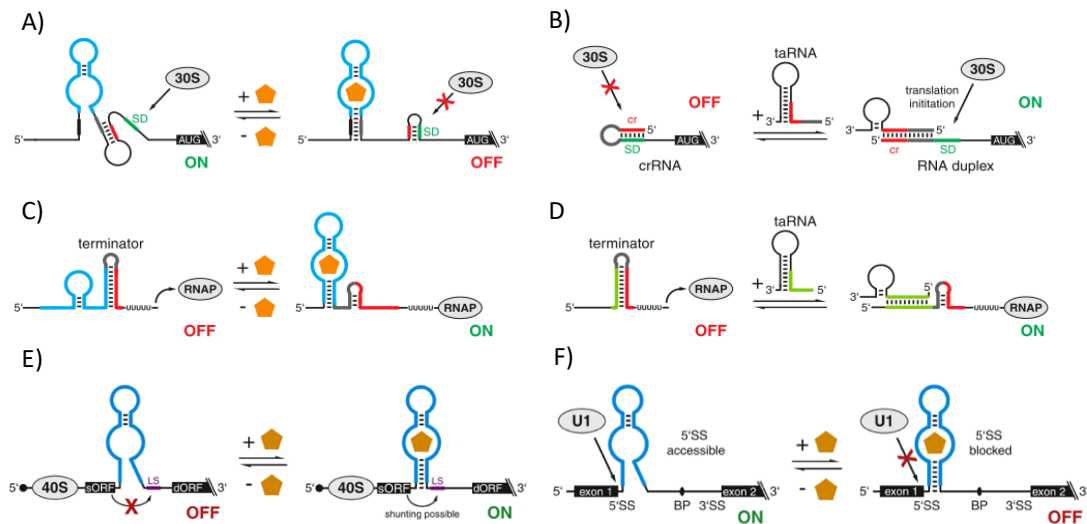


Figure 2. Mechanism of action of several engineered riboswitches.

(Modified from Groher, Süß, 2014) Mechanisms of action in bacteria for engineered riboswitches in prokaryotes (A-D) and eukaryotes (E, F). 30s represent the ribosomal subunit in prokaryotes, SD the Shine-Dalgarno sequence, where the ribosome binds, and taRNA stands for trans-activating RNA. RNAP represent RNA polymerase. 40s is the small ribosomal subunit in eukaryotes, U1 represents the U1 small ribonucleoprotein, sORF for short open reading frame, and dORF for downstream open reading frame. A) represents a regulation of translation initiation via the blocking of the SD sequence B) regulation of translation initiation via taRNA freeing the SD sequence, C the regulation of transcription termination by having the aptamer bind to a complementary sequence); and D), regulation of transcription termination via taRNA. In the case of prokaryotes, E) shows the regulation via ribosomal shunting, where an aptamer domain is inserted between a sORF and a dORF, and ligand binding allows the shunting over to the landing site (LS), and reinitiation of translation; and F) is a diagram for regulation of pre-mRNA splicing, via the integration of an aptamer into the intron of a eukaryotic mRNA, controlling accessibility of splice elements, like the 5' splice site (5'SS), the branch point (BP) or the 3' splice site (3'SS). Ligand binding can either inhibit splicing or enhance exon skipping

### 1.3 Riboswitch improvement: Build-a-switch

Riboswitches are great targets for directed evolution due to their orthogonality, variation, and simplicity, as sequence changes correlate directly to changes in

functionality. There are currently two main ways to improve or develop riboswitches *de novo*: Computer model predictions or Directed Evolution.

The first method requires a previously known aptamer as a working base, and to choose a specific type of model to work with, which could be organised based on the type of data they use for their predictions. Many models use thermodynamics and free energy, accounting for what structures would be more plausible under certain conditions based on mathematical and computational models (Retwitzer, Kifer, Sengupta, & Yakhini, 2015), others use physics and folding mechanisms (Borujeni, Mishler, Wang, Huso, & Salis, 2015), while others use RNA and molecular kinetics for their predictions (Beisel & Smolke, 2009; Gong, Wang, Wang, & Zhang, 2017).

This creates predictions for “ideal” riboswitches fitting the criteria, which would then be synthesised and tested. But herein lies one of the problems of this approach. Computer models are, by definition, incomplete, and also at odds with one another more than once. An ideal riboswitch for model A may not be selected in model B, or the same sequence may have different predicted structures (Barsacchi, Novoa, Kellis, & Bechini, 2016). Even if models agree, the selected molecule may not work at the *in vivo* level, since all the processes happening on a living organism were not part of the simulation. They are, however, very useful, as they give insight into riboswitch structure and functionality, and can be updated in order to take in new constraints as research on the subject develops.

In our laboratory we have developed validated methodologies for the design and prediction of RNA interaction circuits *in vivo* based on the thermodynamics approach (Rodrigo, Landrain, & Jaramillo, 2012; Shen et al., 2015), to the point of engineering a synthetic riboregulator capable of *trans*-activating a *cis*-repressed gene; and a computer model for novel riboswitch development (Rodrigo & Jaramillo, 2014).

#### 1.4 Directed evolution: Playing God on a smaller scale.

The second research line within synthetic biology that can be used for riboswitch development is known as “Directed evolution” (D.E.). The main concept to grasp is a sped-up evolution process towards a specific goal. Experiments are designed around the mutation of proteins and genetic sequences in order to evolve and reach a desired goal (Chen & Romesberg, 2014; Cobb, Sun, & Zhao, 2013). The element of interest undergoes multiple rounds of mutation, selection, and amplification, all in order to progressively obtain refined sequences that would be closer to the desired function than their parental molecules.

In the case of novel riboswitch development, D.E. has become a significant player in the field, being used to detect new aptamers binding to a specific molecule of interest when none are known. This process is called SELEX (Systematic Evolution of Ligands by Exponential enrichment) (Stoltenburg, Reinemann, & Strehlitz, 2007). The rationale behind the method is to select or “fish” molecule-reacting sequences out of a random oligonucleotide library bound to a multi-well plate, in a high-throughput manner. Non-reacting sequences are eliminated from the system, while the ones that do bind to the molecule of interest are amplified and go through increasingly stringent iterations of the same process. The final result is a highly enriched pool containing the selected oligonucleotides, which can then be used to assemble riboswitches.

However, SELEX is not without its caveats (Lakhin, Tarantul, & Gening, 2013; H. Liu & Yu, 2018). To begin with, not every molecule of interest can have an aptamer developed for it, which reduces the possible target selection; and also the sheer amount of random sequences or “sequence space” is massive, and the process needs to be overseen by a researcher at several points, which means it is laborious and time consuming. Despite all these, one of the biggest issue SELEX faces is the fact that it is an *in vitro* process, meaning that even if some molecule-binding aptamers are actually obtained by the end of the process, they might not work properly in a cell environment. Advances have been taken to deal with this issue, developing procedures for *in vivo* SELEX (Mi et al., 2010; H. Wang et al., 2018), but even in those cases, the procedure is cumbersome and slow, relying on

specific modifications in the RNA for stability, and requiring extra steps of RNA purification and PCR in each cycle.

### 1.5 Bacteriophages: Nature's most efficient killing machines

Bacteriophages are a group of virus which predate on hosts within the bacteria and archaea groups, and represent most ubiquitous and varied group among all organisms on Earth (Hendrix, 2003). First hypothesized at the end of the 19<sup>th</sup> Century (Hankin, 1896), and discovered independently twice (D'Herelles, 1949; Twort, 1915), there are currently 19 recognized families within the group (Ackerman, 2009), based on structure and different genome organizations. Due to their variation, sometimes classification of a particular species is a complicated issue, with no efficient answer available (Gibbs, 2013). The use of bacteriophages in both applied and basic research fields cover a wide array of applications. These include phage display (Smith, 1985; Smith & Petrenko, 1997), a procedure for peptide isolation based on their expression on the phage's capsid and their binding to a specific immobilised molecule, which would then be used for their purification. Phages have also recently been incorporated as an element used to ensure Food safety (Moye, Woolston, & Sulakvelidze, 2018), essentially using phages as bio-controlling agents to avoid foodborne illness.

Phages are also a major player in the medical research field, being very relevant for the discovery of possible antibiotic compounds (J. Liu et al., 2004), their use in diagnostics (Schofield, Sharp, & Westwater, 2012), or one of the most impactful lines of phage research nowadays, phage therapy. Phage therapies were first started at the beginning of the 20<sup>th</sup> Century in the Soviet Union (Bunting, 1997; Kuchment, 2012), but fell into disuse in Western countries due to, among other things, the rapid rise of antibiotics, lack of understanding of the subject, and international disinterest in the subject (Kutter et al., 2010). The large scale of the current Antibiotic Crisis has once again brought Phage Therapy into the spotlight (Golkar, Bagasra, & Gene Pace, 2014), and it has been progressively gaining more support and proving useful as a *bona fide* therapeutic agent (R. Brown, Lengeling, & Wang, 2017; B. K. Chan et al., 2018). The final type of research involving phages



are evolutionary experiments (Fernandez-Gacio, Uguen, & Fastrez, 2003), which take advantage of the phage's properties for the directed evolution of biomolecules.

### 1.6 The T7 Phage: You've heard of him.

One of the most well-known phages is the T7 phage, a member of the T4 family of bacteriophages discovered in the early 20<sup>th</sup> Century (Delbrück, 1946; Demerec & Fano, 1945). T7 is an obligate lytic phage infecting *E. coli*, with a genome of roughly 40 kb, coding for 19 different genes (Dunn, Studier, & Gottesman, 1983; Studier & Dunn, 1982), which have made it a very useful tool in studies of genome economy (L. Y. Chan, Kosuri, & Endy, 2005). Its infective properties, such as a short injection and lytic replication time, along with its high replication efficiency (approx. 100 offspring per replication) (Molineux, 2005) and the aforementioned small genome, make it ideal for use in evolutionary experiments (Springman, Keller, Molineux, & Bull, 2010). To further support this, a higher mutation rate can be incorporated by using an error prone copy of the phage's own DNA polymerase (Friedberg, Fischhaber, & Kisker, 2001).

### 1.7 Phage-assisted strategies.

T7 was recently used in the renowned evolutionary strategy known as Phage Assisted Continuous Evolution (PACE) (Esvelt, Carlson, & Liu, 2011). As previously stated, current riboswitch-developing technologies suffer from a common issue, the fact that they do not take place *in vivo*, and this system holds many possibilities for the field of riboswitch directed evolution

The high mutation rates and relatively easy manipulation of T7 bacteriophages, along with the newer technologies of bacterial continuous culture, allow for a more autonomous process of directed evolution that does not require as much time and regulation from the researcher as other *in vitro* processes. The idea at the core of this process is that phages lacking a specific, essential gene are used to infect cells that provide such a gene *in trans*, but with its production regulated



by the desired evolutionary target, which is contained within the phage. A basic schematic of the process can be seen in Figure 3, indicating the way in which the phage drives the evolutionary process, keeping mutations away from the bacterial strains used.

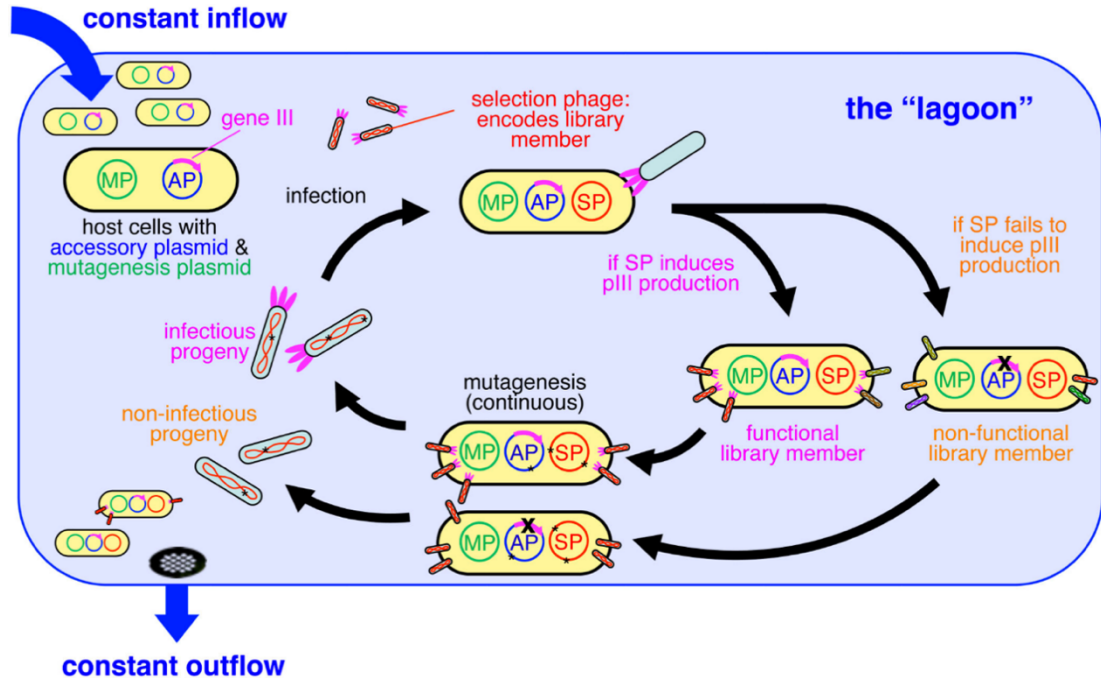


Figure 3. Overview of PACE process.

(From Esvelt, Carlson and Liu, 2011) Schematic of the PACE process, showcasing the most important elements (library phages and host cells with the accessory and mutagenic plasmids), as well as the dynamics between them. Members of the library that do not produce working phages get washed out of the system due to the constant flow of cells, while surviving phages get progressively enriched in the phage lagoon.

This means that only phages that are able to trigger the production of the gene *in trans* by having a working version of its regulator will complete their cycle, survive, and go on to the next step of the process. As the process undergoes subsequent iterations, the desired element will progressively become more functionally efficient, due to the stringency of the evolutionary process. The main issue of the process relies on the complexity of its bioreactor setup, which can represent quite a hurdle initially in order to set it up properly for the correct functioning of the experiments.

In recent years, several developments have taken place that take directed evolution via phages further (Yosef, Goren, Globus, Molshanski-Mor, & Qimron, 2017b). This particular process started with the assembly of multiple plasmids containing leg genes from different phage species, and then transforming them into hosts capable of properly regulating their expression. Infecting these hosts with WT T7 phages produced new phages expressing the different phage legs, which conferred them novel host specificity. This process, based around transduction rather than infection, allowed for phages to act as an efficient carrier system for DNA that was progressively optimised via directed evolution. A schematic of the process can be seen on Figure 4, indicating the main elements and steps of the procedure.

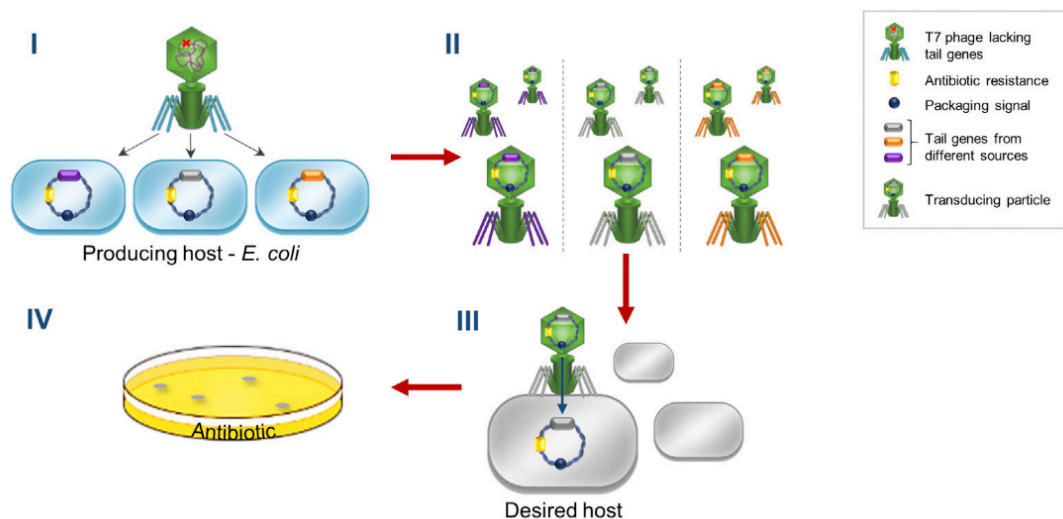


Figure 4. Overview of transduction process to extend phage host range.

(From Yosef et al., 2017b) Schematic of the original transduction process. I) T7 phages are used to infect *E. coli* cell strains carrying specific genes encoding for different phage legs (responsible for their tropism). II) Via transduction, some phages incorporate the plasmids within their capsid, becoming phagemids, and produce the new legs. III) The new phages are used to infect different cells not suitable for T7 replication. If they expressed the proper legs, they would be able to enter the cells. IV) If phages were able to recognize the specific host and introduce the plasmid they carried, those cells would then be able to grow in a plate with specific antibiotic resistance, and the transduction calculated. This system was then used to introduce randomized phage leg genes, which would be optimised via directed evolution.

In my current laboratory, our own continuous culture scheme based on PACE was engineered by designing and developing our own bioreactor, allowing us to directly evolve bacteriophages containing our sequence of interest. This was done by infecting specifically designed selection cells strains in order to select for more progressively efficient activities. Our bioreactor could alternate the inflow of cells between different selection strains in an autonomous way, while maintaining an outflow that ensured cells were not able to replicate within the system; all allowing for just the minimal overseeing by the researcher while experiments were in process. However, in the cases of a malfunction by the bioreactor, the process could be easily done by hand using traditional methods, albeit less autonomously, more akin to the Phage Assisted Non-Continuous Evolution (PANCE) procedure from Suzuki *et al.* (2017).

### 1.8 My project

My research goal was to join both riboswitch development and evolution with directed evolution via phages, in order to use the continuous culture of bacteria to directly evolve a genetic sequence via bacteriophages; which in this case was the previously mentioned T7 virus.

The key to this evolution process lies within phages containing a riboswitch sequence of interest; but the process could also be used for the evolution of other types of molecules as well. During the course of this project, there was an effort to evolve protein receptors using this same strategy, as part of a collaborative effort within the Life Sciences department at the University of Warwick. In order to do this, we took advantage of previously developed methodologies, assembling synthetic riboregulators based on our predictions, and then inserted them in the genome of T7 phages via homologous recombination (abbreviated HR in tables) (Pires, Cleto, Sillankorva, Azeredo, & Lu, 2016), a natural system by which 2 distinct DNA sequences flanked by the same homologous regions exchange location. By designing specific plasmids containing regions homologous to the phage's genome, we could introduce our sequences of interest and at the same

time eliminate specific genes from the phage, allowing us to control it by expressing those genes in *trans*.

In order to select these phages, the homologous recombination process exchanged the T7 phage's *gene 5*, encoding its DNA polymerase (Molineux, 2005); for our gene fragment of interest followed by the gene *trxA*, which encodes the protein thioredoxin A, an *E. coli*-produced protein essential for the phage's replication cycle, as it works as the processivity factor for the phages DNA polymerase (Qimron, Marintcheva, Tabor, & Richardson, 2006). A deletion of this gene will directly affect the phage growth, which would impact the phage population's "Efficiency of plating" (EOP), defined by Adams as "the plaque titre of a phage preparation, determined by plating under stated conditions, relative to some other estimate, usually higher, of the concentration of phage particles" (Adams, 1959).

In the case of the directed evolution of a riboswitch, the recombined gene fragment contains a T7 promoter, the theophylline riboswitch (Lynch, Desai, Sajja, & Gallivan, 2007) controlling the expression of a repressor originating from phage  $\lambda$ , known as the *cI* repressor (Fehér, Karcagi, Blattner, & Pósfai, 2012); and essential gene *trxA*. The *cI* repressor regulates the switch from a lysogenic state in the case of phage  $\lambda$ , maintaining latency and preventing DNA replication and the exit of the phage (Dodd, Perkins, Tsemitsidis, & Egan, 2001).

The rationale behind the use of the phage  $\lambda$  *cI* repressor was that it acts as a transcriptional activator, binding to a specific promoter known as the pRM promoter (D. Huang, Holtz, & Maharbiz, 2012); and so, its conditional expression thanks to the riboswitch could be used to regulate that of subsequent genes, using it as a dam of sorts to control the selection process.

A representation of the fragment can be seen in Figure 5, as well as the homology regions and *gene 5* in the phage's genome.

There were two phage populations used during these experiments. One contained an already established working version of the riboswitch (from Lynch et al. 2007) which will be used as a control; while the other is a phage library of riboswitches containing 8 random nucleotides at the end of the stem (section marked “N8” next to the riboswitches in Figure 5). This created a prospective library of  $6.5536 \times 10^4$  different riboswitches to begin the evolutionary process with. Over the course of the rounds of selection, the number of variants within the population was expected to decrease, making for more homogeneous libraries. Initially, the ones that did not work due to basic structural constraints would be eliminated, and as the selection went on, conditions could be modified to increase its stringency, such as using a lower concentration of theophylline to select variants with higher binding rates. Information regarding the production of these libraries can be found in Chapter 3.

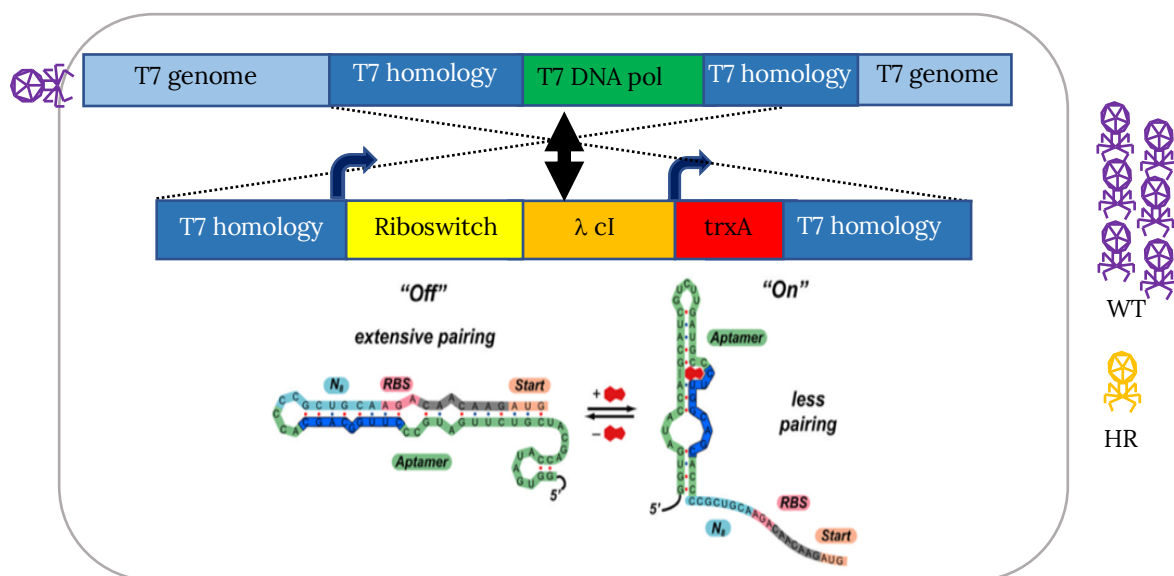


Figure 5. Homologous recombination of the sequence of interest into T7 phages.

General scheme of T7 homologous recombination, showing the T7 homology regions flanking the fragments of interest, and the T7 DNAPol, which the phage will lose in exchange for it. Out of the phage’s offspring, only a small percentage of them will undergo this procedure.

The recombinant phages would undergo 2 different continuous culture conditions; a positive selection, where the molecule of interest was present and thus the riboswitch selected for an active/ “ON” state, and a negative selection where the molecule of interest wasn’t present, and so the selected riboswitch’s

state should be inactive/ “OFF”. In both selections, in order to find a suitable candidate that shows a significant activation fold, multiple combinations of the pRM promoter and different Ribosome Binding Sites (RBSs) of variable strength had to be tested. Five of these RBSs originate from the iGEM registry of standard biological parts (<http://parts.igem.org>, first described in Weiss 2001) and one was obtained via the use of the RBS calculator (Borujeni, Channarasappa, & Salis, 2013). Both strains also contain the mutation-inducer compound N-methyl-N'-nitro-N-nitrosoguanidine (NG) to further increase the mutation ratio of the system (Sugimura, Nagao, & Okada, 1966), as well as an error-prone version of the phage's DNA polymerase.

The experimental setup was designed in such a way that cells were flowing in and getting washed out of the system before they could replicate, but however remained for enough time for the phages to do so. That way, any mutations that can affect the evolutionary process would only be conserved within the phage population. All details about this procedure can be found in Chapter 4.

These findings could then be used to improve current riboswitch prediction models (Gong et al., 2017) by feeding these new parameters into them. Obviously, as previously mentioned when regarding any model, there is no perfect one, since there are multiple elements not taken into account, due to the difficulties that would arise from including too many in the same model, or simply because the supported theory does not contemplate them. One of the biggest challenges would be the incorporation of the *in vivo* data into models, something for which this research could be helpful. By analysing the resulting riboswitch structures after the evolutionary procedure, conclusions could be drawn based on the different nucleotides present, not only from a structural, but from a functional standpoint as well; which could be included in some of the aforementioned models in order to improve their prediction power.

The presented experimental setup was also valid for other types of sensors that are trying to be evolved, with the main elements driving the selection being the same in all cases. One such alternative evolution is the case of the MmfR/MmyB

system, regulating biosynthesis of the molecules known as methylenomycin furans (MMF) (O'Rourke et al., 2009), which was tackled as part of a collaboration with Dr Christophe Corre at the University of Warwick.

The *mmyB* gene codes for a protein regulating the synthesis of MMF, and in order to regulate the system, once the concentration is high enough, the MMF molecules bind to the MmfR protein encoded in gene *mmfR*, and repress the action of MmyB. All details about this procedure can be found in Chapter 4.

In recent years, as explained previously, my laboratory has also started developing a new strategy discussed in the literature (Yosef et al., 2017b) based upon the use of “phagemids”; bacterial plasmids containing a sequence of interest as well as a phage’s packaging signal and an antibiotic resistance, in order for some of the viral offspring to pack it within their virion during the replication process (Chung & Hinkle, 1990b). By keeping the riboswitches in a plasmid library, rather than recombining them into the phages, a higher activation could be achieved, by growing cells with theophylline in the media for enough time to fully activate any working variants present. And by eliminating the full phage in favour of phagemids, the replication only happens under very specific terms, and can be triggered when needed, rather than having to accommodate everything in the original seventeen minutes of T7 replication. All details of this procedure can be found in Chapter 5.

As a side project throughout the final stretches of this PhD, a different project was undertaken, which, however, kept similar beats to the previous work that conforms this thesis. The development of an inducible gene expression system based upon the genome of the Q $\beta$  bacteriophage (Gorzelnik et al., 2016). This (+) ssRNA phage has a very compact genome with only 4 genes, one of which is a well-conserved RNA-dependent RNA polymerase known as replicase (D. Brown & Gold, 1996; Kashiwagi & Yomo, 2011). By eliminating one of the phage’s genes and substituting it for an antibiotic resistance gene, experiments were designed to confirm whether this would be a viable novel system for transient and conditional gene expression to be developed. All details regarding this procedure can be found in Chapter 6.



## AIMS AND MAIN OBJECTIVES

RNA regulation is an essential part of gene expression in living organisms, and a key player in the advance of synthetic biology systems. The aims of this thesis revolve around developing efficient ways in which phages can be harnessed to develop new ways to use and improve current RNA regulation systems, as well as to create *de novo* elements for multiple research settings.

To summarise, the main objective of this thesis was as follows:

- To set up a system for the directed evolution of riboswitches using T7 phages in a continuous culture scheme, via the development of a phage riboswitch library and specific selection strains.

As the project moved forward, secondary objectives arose which were directly related to the main one:

- To set up an alternative strategy using phagemids instead of full phages.
- To test these systems in the evolution of alternative molecules, such as cell membrane receptors.

In one case, separated from but following a similar path to all previous experiments, the objective was:

- To develop a transient RNA-dependent system for the tuneable expression of genes in bacteria by incorporating them on the genome of bacteriophage Q $\beta$



## CHAPTER 2. MATERIALS AND METHODS

### 2.1 Original cell strains

*E. coli* DH5 $\alpha$ , TOP10, MG1655Z1, from the laboratory's own stock; and cells from the Keio Collection (Baba et al., 2006), namely the parental Strain *E. coli* BW25113, and KO strains for essential phage genes *cmk*, *trxA*, and both (CGSC, <http://cgsc2.biology.yale.edu>). The specifics for each of these strains can be found in , and all are available in the laboratory's LIMS.

Table 1. List of *E. coli* strains used throughout the PhD.

List of cell strains used throughout the PhD as part of all the different projects. Keio collection strains described in Baba et al., (2006) and found at the Coli Genetic Stock Center (<https://cgsc2.biology.yale.edu/KeioList.php>)

Strain name	Genotype	Project & Origin
<i>E. coli</i> DH5 $\alpha$	F <sup>-</sup> <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20</i> $\phi$ 80 <i>dlacZ</i> $\Delta$ M15 $\Delta$ ( <i>lacZYA-argF</i> )U169, <i>hsdR17</i> ( <i>r<sub>K</sub><sup>-</sup>m<sub>K</sub><sup>+</sup></i> ), $\lambda^-$	P) Initial full phage strategy (Discarded) O) Lab stocks
<i>E. coli</i> TOP10	F <sup>-</sup> <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) $\phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74 nupG recA1 araD139</i> $\Delta$ ( <i>ara-leu</i> )7697 <i>galE15 galK16 rpsL</i> (Str <sup>R</sup> ) <i>endA1</i> $\lambda^-$	P) Initial full phage strategy (Discarded), Phagemid strategy O) Lab stocks
<i>E. coli</i> MG1655	K-12 F <sup>-</sup> $\lambda^-$ <i>ilvG<sup>-</sup> rfb-50 rph-1</i>	P) Initial full phage strategy, (Discarded) O) Lab stocks
<i>E. coli</i> BW25113	<i>lacI<sup>+</sup>rrnB<sub>T14</sub></i> $\Delta$ <i>lacZ<sub>WJ16</sub></i> <i>hsdR514</i> $\Delta$ <i>araBAD<sub>AH33</sub></i> $\Delta$ <i>rhaBAD<sub>LD78</sub></i> <i>rph-1</i> $\Delta$ ( <i>araB-D</i> )567 $\Delta$ ( <i>rhaD-B</i> )568 $\Delta$ <i>lacZ4787</i> (:: <i>rrnB-3</i> ) <i>hsdR514 rph-1</i>	P) Full phage strategy, cloning procedures, Phagemid strategy O) Lab stocks, SAJ128
<i>E. coli</i> BW25113 $\Delta$ <i>trxA</i> (Keio collection JW5856)	F <sup>-</sup> , $\Delta$ ( <i>araD-araB</i> )567, $\Delta$ <i>lacZ4787</i> (:: <i>rrnB-3</i> ), $\lambda^-$ , <i>rph-1</i> , $\Delta$ <i>trxA732::kan</i> , $\Delta$ ( <i>rhaD-rhaB</i> )568, <i>hsdR514</i>	P) Full phage strategy, phagemid strategy O) Lab stocks, SAJ130

E. coli BW25113 $\Delta cmk$ (Keio collection JW0893)	F <sup>-</sup> , $\Delta(araD-araB)567$ , $\Delta lacZ4787(::rrnB-3)$ , $\lambda^- \Delta cmk-734::kan$ , $rph-1 \Delta(rhaD-rhaB)568$ , $hsdR514$	P) Full phage strategy, phagemid strategy O) Lab stocks
E. coli BW25113 $\Delta trxA \Delta cmk$	F <sup>-</sup> , $\Delta(araD-araB)567$ , $\Delta lacZ4787(::rrnB-3)$ , $\lambda^- \Delta trxA732::tet$ , $\Delta cmk-734::kan$ , $rph-1 \Delta(rhaD-rhaB)568$ , $hsdR514$	P) Full phage strategy, phagemid strategy O) Lab stocks, SAJ19
E. coli LEMO21 DE3	$fhuA2$ [lon] $ompT$ $gal$ ( $\lambda$ DE3) [dcm] $\Delta hsdS$ / pLemo(Cam <sup>R</sup> )	P) Q $\beta$ strategy O) New England Biolabs
E. coli BL21 DE3	E. coli str. B F <sup>-</sup> $ompT$ $gal$ $dcm$ $lon$ $hsdS_B(r_B^- m_B^-)$ $\lambda(DE3$ [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB <sup>+</sup> ] <sub>K-12</sub> ( $\lambda^S$ )	P) Q $\beta$ strategy O) New England Biolabs, SAJ733
E. coli BL21 AI	E. coli B F <sup>-</sup> $ompT$ $gal$ $dcm$ $lon$ $hsdS_B(r_B^- m_B^-)$ [malB <sup>+</sup> ] <sub>K-12</sub> ( $\lambda^S$ ) $araB::T7RNAP-tetA$	P) Q $\beta$ strategy O) Invitrogen, SAJ760

## 2.2 Original phage strains

All phages used throughout this PhD were T7 phages, part of the laboratory stocks. Phage Q $\beta$  is not included as part of this list due to its genome being sequenced inside a custom-ordered plasmid, not as a free phage.

Table 2. List of original T7 strains used throughout the PhD

T7 strains and the projects they were used for during the length of the PhD.

Strain	Project	Origin
WT T7	Full phage strategy, phagemid strategy	Lab stocks
T7 $\Delta gp5$	Phagemid strategy	Lab stocks
T7 $\Delta gp5$ $trxA^+$	Phagemid strategy	Lab stocks

## 2.3 Lysogeny broth media

All bacterial strains were grown in LB unless specified accordingly, with any alternative media used only for specific experiments. For 1 L of broth: 1 L H<sub>2</sub>O, 10 g Peptone 140, 5 g yeast extract, and 5 g sodium chloride are needed.

## 2.4 Theophylline solution

In order to obtain a stock solution of 25 mM, 225 mg of powdered theophylline were diluted in 50 ml of dH<sub>2</sub>O and vortexed until fully diluted, then aliquoted into 1.5 ml Eppendorf tubes in order to avoid thawing more than the needed quantity for the required experiments at any given time.

## 2.5 Saline-Magnesium (SM) buffer

Solution used for long-term storage of phages at 4°C, stabilizes them and avoids the risk of contamination by microorganisms. For 1L of SM buffer: 1 L water, 50 mM Tris-HCl 7.5, 100 mM NaCl, and 100 mM MgSO<sub>4</sub>. (Based upon the same recipe used in Lee, Billington, Hudson, & Heinemann, 2011)

## 2.6 M9 media

M9 is a minimal media used for fluorometric assays, to measure the expression of GFP. For a total of 200 ml, 100 uM of CaCl<sub>2</sub>, 2 mM of MgSO<sub>4</sub>, 10 uM of FeSO<sub>4</sub>, 40 ml of 5X M9 salts, 3.2 ml of 50% glycerol, 8 ml of 5 % cas-amino acids, 1 µg/ml of thiamine, 20 mg/ml of uracil, 30 µg/ml of Leucine, 50 µL of NaOH 8 M and add until pH 7.4; and finally complement to 200 ml with water. (Anonymous, 2010)

## 2.7 Antibiotics

Throughout all experiments described in this thesis, different antibiotics have been used in growth media or plates for bacterial culture and are indicated when appropriate. Here are the names and stock concentrations of the most used ones.

Table 3. List of antibiotics used throughout the PhD.

Ampicillin (Amp)	Chloramphenicol (Cm)	Kanamycin (Kan)	Gentamycin (Gent)	Tetracycline (Tet)
100 mg/ml 0.286 M	35 mg/ml 0.108 M	50 mg/ml 0.103 M	10 mg/ml 0.021 M	10 mg/m 0.022 M

All cultures and plates were made with a 1/1000 dilution of the corresponding antibiotic or antibiotic mix. E.g. 5 ml of an Amp culture would have 5 µl of Amp added to it, while an Amp/Kan would have 2.5 µl of each to keep the same ratio.

### 2.8 Competent cell production

For transformation purposes, many of the cell lines used needed to be made competent at different stages of the project, so two methods were used, one to make electrocompetent cells, and another to make chemically competent ones.

#### 2.8.1 Electrocompetent cells:

Prepared according to the Krantz Lab protocol ([http://mcb.berkeley.edu/labs/krantz/protocols/electrocomp\\_cells.pdf](http://mcb.berkeley.edu/labs/krantz/protocols/electrocomp_cells.pdf)).

10 ml of overnight culture was added to 1 L of LB and grown at 37°C with shaking until the OD<sub>600 nm</sub> reached exponential culture, between 0.35 and 0.4. These exponentially growing cells were kept at 4°C for the remainder of the process. The volume was then split into ice-cold centrifuge tubes and the cells pelleted at 1000 *g* for 20 min at 4°C. The supernatant was discarded, and the pellets resuspend in ice-cold sterile water, dividing the number of tubes in half by combining the pellets. This procedure was repeated twice more, each time halving the number of tubes and the volume of water. After the last pelleting of the cells, the cells were resuspended in ice-cold 10% glycerol, halving the volume once more. The cells were pelleted, and the supernatant discarded, each pellet was resuspended in 1 ml of ice-cold glycerol. 50 µL of the suspension were aliquoted into chilled Eppendorf tubes and stored at -80°C.

#### 2.8.2 Chemically competent:

10 ml of an overnight culture was added to 1 L of LB and grown at 37°C with shaking, until OD<sub>600</sub> was between 0.35 and 0.4. The cells were then put on ice for 30 minutes, before being pelleted by centrifugation (3000 *g*, 4°C for 5 minutes). The supernatant was discarded, and the cells resuspended in 10 ml of 100 mM CaCl<sub>2</sub>. The cells were incubated on ice for 30 min, before the cells pelleted by

centrifugation using the same conditions as previously described. The supernatant was again discarded, and the cells resuspended in 10 ml of 100 mM  $\text{MgCl}_2$  and incubated on ice 30 minutes. Finally, they were pelleted again, the supernatant discarded and resuspended into 2 ml of  $\text{CaCl}_2$ +10% glycerol, ensuring thorough mixing. 50  $\mu\text{L}$  aliquots were stored at  $-80^\circ\text{C}$  for use in transformations.

### 2.9 Transformations

Plasmids were transformed into both electrically and chemically competent cells. In the first case, a minimum of 100 ng of DNA were added to 50  $\mu\text{L}$  of competent cells and left on ice for 15 minutes, along with a 2-mm gap electroporation cuvette from BioRad (BioRad #1652086). The volume was then transferred to the cuvette, the cells electroporated using a MicroPulser Electroporator from BioRad (BioRad #1652100) and then grown in LB media for 1 hour, before being plated. For chemically competent cells, the DNA was added to the 50  $\mu\text{L}$  of cells, which were left on ice for 30 minutes, then heat-shocked at  $42^\circ\text{C}$  for 90 seconds, and put back on ice for 10 minutes, before recuperating them in LB media for an hour. Cells were then plated on a LB agar plate with antibiotics corresponding to the resistance carried by the plasmid.

### 2.10 Plasmids

All plasmids are presented with its original nomenclature and the specific one used throughout the project. Presented here are the parental plasmids used as template for all others throughout the project. The assemblies were made either by enzymatic digestion at  $37^\circ\text{C}$  and overnight ligation at  $20^\circ\text{C}$ , or using the process known as Golden Gate ligation (Engler, Kandzia, & Marillonnet, 2008), where indicated as such. The details on each specific variation based on these plasmids will be discussed in the relevant chapters. All plasmid maps can be found as part of the Supplementary Materials section.

Table 4. List of parental plasmids used throughout the PhD.

Name	Figure	Size	Parental plasmid	Contents
Pr100 pLit Chlor/EG_001	Figure 6	4226 bp ( <i>cmk</i> ) 3490 bp ( $\Delta cmk$ )	pLITMUS 28 (NEB#3528)	pMB1 Ori, Cm <sup>+</sup> . T7 homology regions 4.7 & 5.3 flanking pT7 followed by <i>cmk</i> , <i>cl</i> , & <i>trxA</i> . <i>cmk</i> eliminated via digestion with BglIII and BamHI
pSEVA631/EG_002	Figure 7	3001 bp	SEVA plasmids (Silva-Rocha et al., 2013),	Default cloning site, flanked by <i>rrnB</i> T1 terminator, $\lambda$ t0 terminator, pBBR1 Ori, Gent <sup>r</sup>
Parental positive selection plasmid/EG_002+	Figure 8	7135 bp	EG_002	Same as EG_002, with pTy $\rightarrow$ EP T7 DNApol, BsaI sites flanking pR0010 and RBS controlling RFP, terminators B0010, B0012, selection gene <i>cmk</i> .
Parental negative selection plasmid/EG_002-	Figure 9	8677 bp	EG_002	Same as EG_002, with pTy $\rightarrow$ EP T7 DNApol, BsaI sites flanking pR0010 and RBS controlling RFP, terminators B0010, B0012, selection gene <i>pifA</i> .

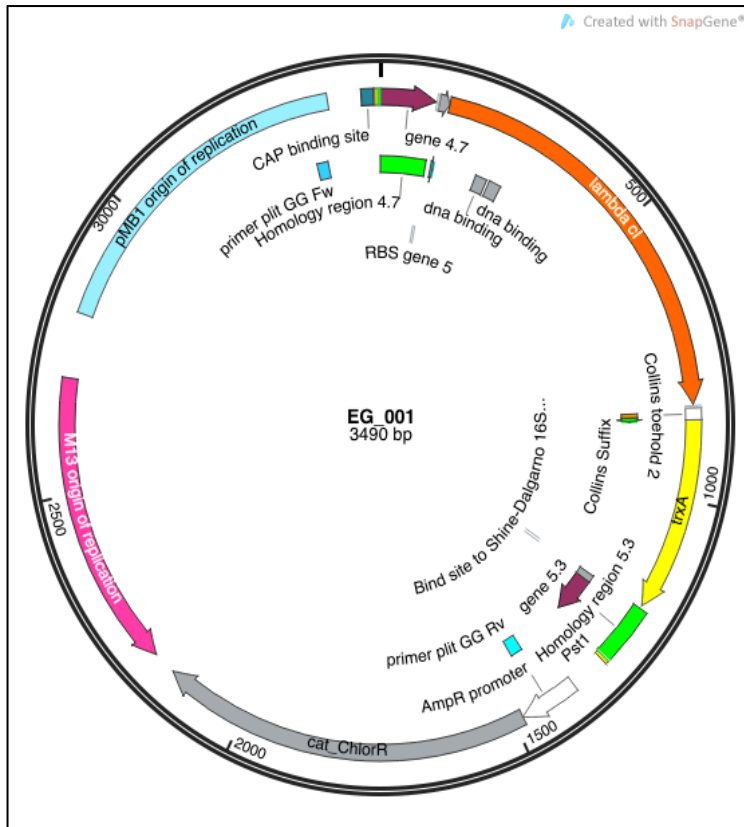


Figure 6. Parental plasmid EG\_001.

Plasmid map of parental homologous recombination plasmid EG\_001, indicating the phage recombination elements 4.7 and 5.3, flanking the lambda cI gene, as well as the *trxA* gene.

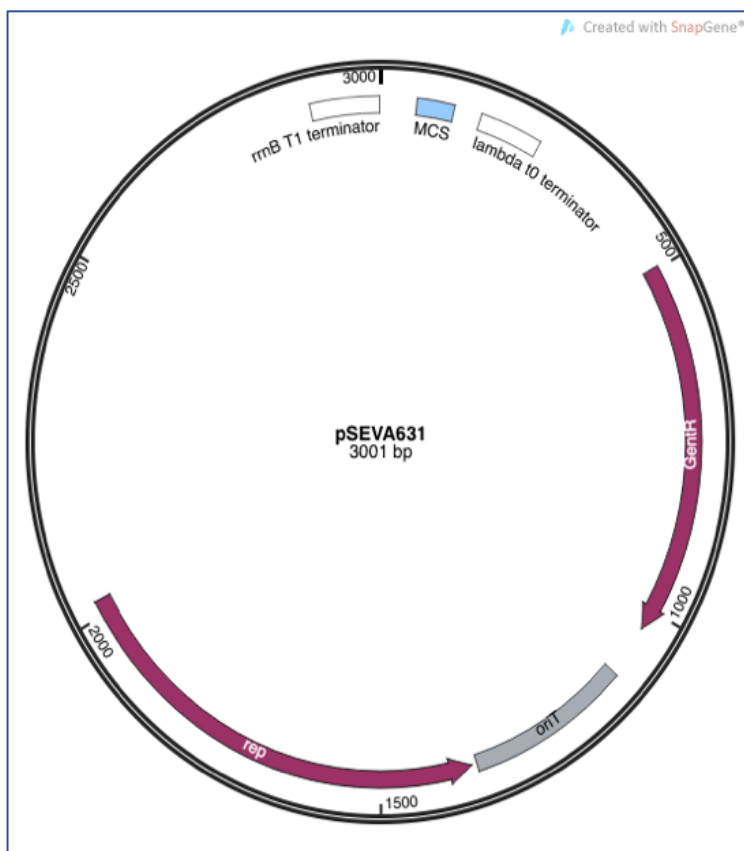


Figure 7. Parental plasmid pSEVA631.

Structure for the pSEVA631 plasmid, used to assemble all other selection plasmids. Based upon the European standard of SEVA plasmids (Silva-Rocha et al., 2013), allowing for easily reproducible modularity.

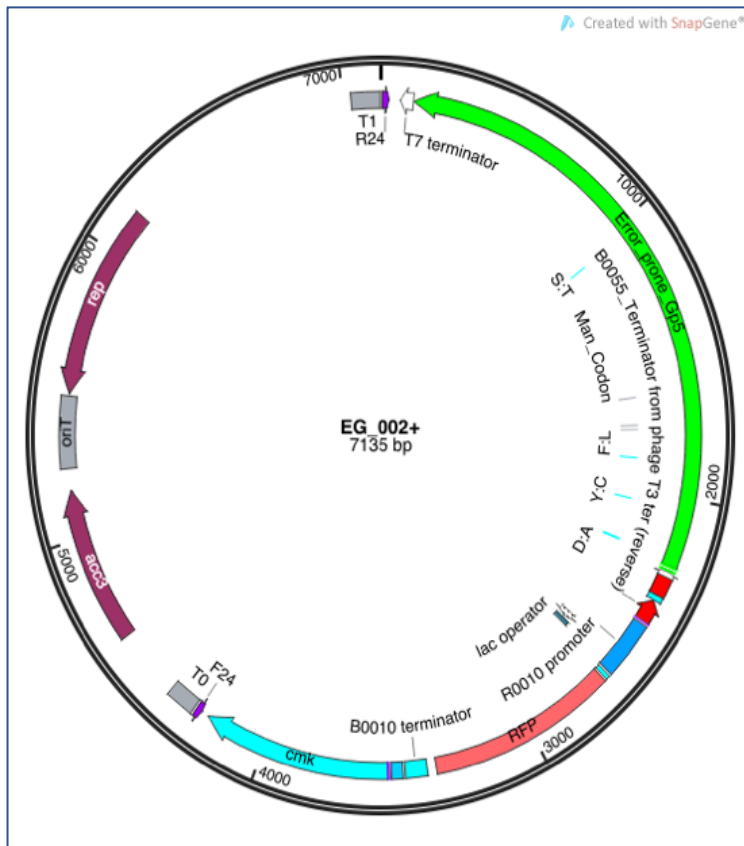


Figure 8. Positive selection parental plasmid EG\_002+. Positive selection parental plasmid, used as a template for all subsequent plasmids used in the procedure, including controls.

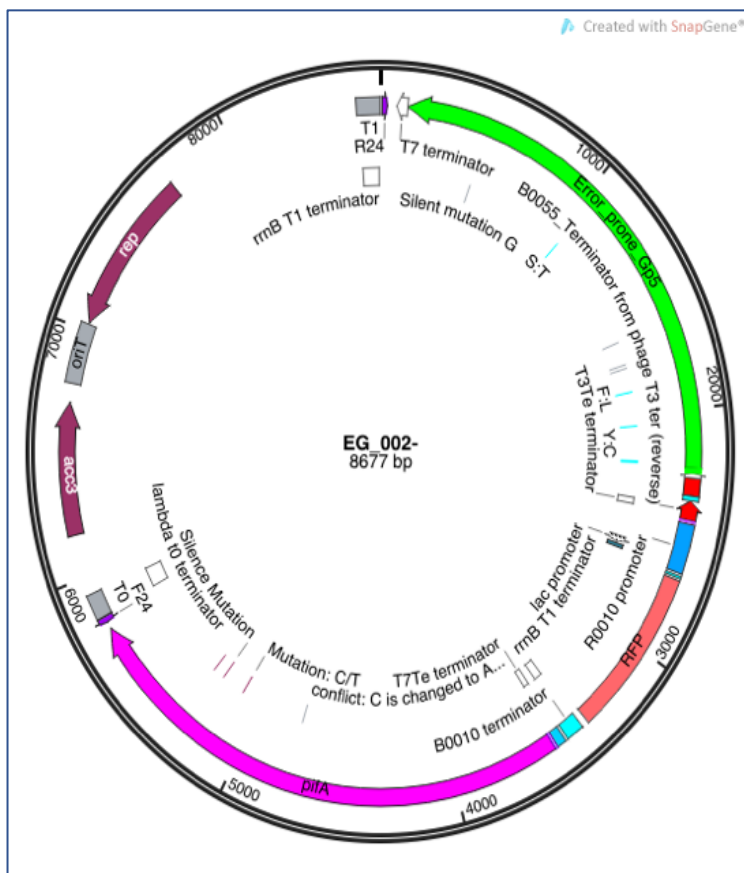


Figure 9. Negative selection parental plasmid EG\_002-. Negative selection parental plasmid used as a template for all negative selection plasmids over the use of this research, including controls.



## 2.11 Primers and Gblocks

All gene sequence constructs produced by Integrated DNA Technologies. Melting temperatures ( $T_m$ ) of the primers were calculated using the NEB calculator ([www.neb.com/external-links/tm-calculator](http://www.neb.com/external-links/tm-calculator)). All primers' name, sequence and designed use can be found in Table 5.

Table 5. List of primers used throughout the thesis.

List of main primers used throughout the different strategies, divided into newly designed primers and pre-existing ones. Specific sub-categories are present in the table. For each primer there is its the name, sequence and specific use.

<u>PRIMER NAME</u>	<u>USE</u>	<u>SEQUENCE</u>
		<b>Designed primers</b>
EG_GG01	HR Goldengate fragment addition Fw	cttcgaagacgataaccagcatcgtcttgatgcctt tggcagcacctgagaccatagaatgagcacaaaa aaggtctcgaaccattaaca
EG_GG02	HR Goldengate fragment addition Rev	cttcgaagacctggtatcaccccctatagagtctg attaacaaaaaacccctaccccccgatatgtatt atcgtgacttaacaatctct
EG_GG03	cI insert + Rd Rbsw Fwd	aaggtctcccacnnnnnnnnagacaacaaga tgagcacaaaaaagaaaccgagaccaa
EG_GG04	cI insert + Fix Rbsw Fwd	aaggtctcccaccccgctgcaagacaacaagatg agcacaaaaaagaaaccgagaccaa
EG_GG05	cI insert+ Rbsw Rev	ttggtctcggtttctttttgtgctcatcttgtgtc
EG_064	Goldengate presence test Fw	tgccgataagaccaacatga
EG_065	Goldengate presence test Rv	gcatcaagacgatgctggta
EG_068	Eliminate <i>cmk</i> from HR plasmid Fw	ataaacagaggagatatcacatgaccgaattgc accggt
EG_069	Eliminate <i>cmk</i> from HR plasmid Rev	agtagcctcctgtagtcgacttattaggccagcgc cagtt
EG_096	Detect <i>pifA</i> presence	cggaagaaagagcgcttcaaccggactgccg
EG_097	Gene 5 Fwd	ctcaaggtcgccaagtgg
EG_098	Gene 5 Rev	gaaggacactttaatttcttgcatagcc
EG_099	Change RBS for <i>cmk</i> in <i>mmv</i> plasmid (Fw)	aggtctcaataagctcatccgcggsaataattgag raggwacagataatgaccgaattgcaccggtaat tacg

EG_100	Change RBS for <i>cmk</i> or <i>pifA</i> in <i>mmyb</i> plasmid (Rev)	aggtctcattattacgctgcaagggcgtaattttcg tcg
EG_101	Change RBS for <i>pifA</i> in <i>mmyb</i> plasmid (Fw)	aggtctcaataacccgacaaacccgactmcataa gswggtagcagtagaagatactgaggcagctgt ggaacc
EG_102	Change origin of replication in <i>mmyb</i> plasmid (Fw)	aggtctcagcacgatatacaggattttgcaaagg gttcg
EG_103	Change origin of replication in <i>mmyb</i> plasmid (Rev)	aggtctcagtcgcgaaaaaggatggatataccgaa aaaatcgc
EG_104	Assembly of control for negative selection (Fw)	gtaaggcgccgctcgtcat
EG_105	Assembly of control for negative selection (Fw)	tcatatgacgacggcgcc
EG_106	Gene 5 base addition for insertion in phagemid plasmid Fw	aggtctcatcgagataaacagaggagatcacat gattgtg
EG_107	Phagemid base addition for <i>gene</i> 5 insertion Rev	aggtctcatcgagtcagccaaacgtctcttc
EG_108	Gene 5 base addition for insertion in phagemid plasmid Rev	aggtctcagtcgacttagtggcaaatcgcc
EG_109	Phagemid base addition for <i>gene</i> 5 insertion Fw	aggtctcacgacctgcagtaaagcccgttcg
EG_112	Bind to middle of <i>pifA</i> for detection	atgatctggaccgactggagcc
EG_115	Confirmation of plasmid presence 1 (Fw)	gggtttttgtttgcaagcagcagattacg
EG_116	Confirmation of plasmid presence 1 (Rev)	cggggagtcaggcaactatgg
EG_117	Confirmation of plasmid presence 2 (Fw)	gtatttagaaaaataacaatataggggtccgcg
EG_118	Confirmation of plasmid presence 2 (Rev)	ttgtgagcggataacaatttcacacagg
EG_119	Confirmation of plasmid presence 3 (Fw)	ctgggggtgcctaataagtgagc
EG_120	Confirmation of plasmid presence 3 (Rev)	gtcgattttgtgatgctcgcagg
EG_123	Mid $\lambda$ cl to check for riboswitch (Rev)	gttgtgcttaccatctctccg

EG_124	Addition of overhangs for incorporation of GFP into plasmid (Fw)	tttagcttccttagctcctgataggtataaccacgttt cgcggttagcc
EG_125	Addition of overhangs for incorporation of GFP into plasmid (Rev)	ggcaaattcccatgcaactcgccagggttgataa acgc
EG_126*	Addition of overhangs to GFP for incorporation into plasmid (Fw)	aagttgcatgggaatttgccacttattgtatagttc atccatgccatgtg
EG_127*	Addition of overhangs to GFP for incorporation into plasmid (Rev)	tcaggagctaagggaagctaaaatgactagccgta aaggagaaga
EG_128	Confirmation of GFP plasmid presence 1 (Fw)	caacgatcaaggcgagttacatgatccc
EG_129	Confirmation of GFP plasmid presence 1 (Rev)	aactatacaaataagtggaattcccatgcaact tcgcc
EG_130	Confirmation of GFP plasmid presence 2 (Fw)	cgcaccctaggttagcaacttaactaaccttct cgaaagagagagt
EG_131	Confirmation of GFP plasmid presence 2 (Rev)	tgttaactcgccttgatcgttggaaccggagctg
EG_132	Confirmation of GFP plasmid presence 3 (Fw)	tcctttacggctagtcattttagcttccttagctcct gatagg
EG_133	Confirmation of GFP plasmid presence 3 (Rev)	gtttgctaacctagggtgcgtatag
<b>NGS procedure</b>		
EG_NGS _Fwd	Initial NGS amplification primer, 2 bases missing, Fw	tcgtcggcagcgtcagatgtgtataagagacgtca cgataatacatatcggggggtagg
EG_NGS _Rev	Initial NGS amplification primer, 2 bases missing, Rev	gtctcgtgggctcggagatgtgtataagagacctt taaggcgacgtgcgtcctc
EG_NGSFw_ Final_Full	NGS amplification primer Fw	tcgtcggcagcgtcagatgtgtataagagacaggt cacgataatacatatcggggggtagg
EG_NGSRev_Final _Full	NGS amplification primer Rev	gtctcgtgggctcggagatgtgtataagagacagc ttaaaggcgacgtgcgtcctc
<b>LABORATORY STOCK PRIMERS</b>		
<b>Negative Selection Assembly</b>		
pifA_T7te _Fw	Addition of sequence and Eco3II site upstream of <i>pifA</i> (Fw)	ttcgggtgggcctttctgcgtttataactagtggtc tccatgaagatactgaggcagctg
re_prm30	Addition of sequence containing RBS B0030 and Eco3II site (Rev)	ttatttggtctctcatctagttttctcctcttaaat ctctagacaagatacattgtaaa

re_prm31	Addition of sequence containing RBS B0031 and Eco31I site (Rev)	ttatttggctcttcatctagtaggttctgtgtga ctctagacaagatacattgtaaa
re_prm32	Addition of sequence containing RBS B0032 and Eco31I site (Rev)	ttatttggctcttcatctagtagtcttctgtgtgact ctagacaagatacattgtaaa
re_prm33	Addition of sequence containing RBS B0033 and Eco31I site (Rev)	ttatttggctcttcatctagtagtctgtgtgactc tagacaagatacattgtaaa
re_prm64	Addition of sequence containing RBS B0064 and Eco31I site (Rev)	ttatttggctcttcatctagtagtattccctctttctc tagacaagatacattgtaaa
re_prmSyn	Addition of sequence containing Synth RBS and Eco31I site (Rev)	ttatttggctcttcatctctacctccttaatgtgtt ctattgacgactcacaagatacattgtaaa
T7rbsFWD	Addition of T7 RBS sequence and Eco31I cut site upstream of <i>pifA</i> (Fw)	aataaaggctctgtgaataatacgcactcactatagg gagtcgtcaatagaacacatt
T7rbsREV	Addition of sequence containing T7 RBS and Eco31I site (Rev)	ttatttggctcttcatctctacctccttaatgtgtt ctattgacgactc
<b>Diagnostic</b>	Binding to:	
seq_pLitGG_Fwd	PmB1 Origin of replication	gtcgattttgtgatgctgcgcag
WR_PR4	Homology region 5.3	ttggctcttgccatactagcatttcagaagcagca cc
WR_PR6	Homology region 5.3	ttggctcttgccacacctgtaagtgtcttcgttc
WR_PR8	Homology region 5.3	ttggctcttgccattcatgagtagcctcctgtagtc g
PS2_T0Rev	T0 region in selection plasmids	gcggcaaccgagcggttc
VF2	G00100_VF2 annealing site in selection plasmids	tgccacctgacgtctaagaa
gp5EP_Rev	Error prone T7 gene 5	cacggttgagttgcagctta
Gp5_mtplx_Fwd	Error prone T7 gene 5	tcaccgacgttggtatactac
Gp5_mtplx_Rev	Error prone T7 gene 5	ccatcctgcttctgcagttt
PS1_T1Fw	T1 region in selection plasmids	agggcggcggattgtcc

## 2.12 Restriction enzymes

All used enzymes were part of the FastDigest line from Thermo Scientific. All digestions were carried out at 37 °C, using 200 ng of DNA template, 1 µl of FastDigest enzymes, and 2 µl of 10x buffer solutions from ThermoFisher in a final volume of 20 µL.

### 2.13 Ligation

A mixture of 100 fmol of the corresponding plasmid backbone and 100 fmol of the desired insert, along with 2 µl of Thermo Scientific's 10x Ligase buffer and 2 µl of T4 ligase (ThermoFisher #EL0014) in a final volume of 20 µl, was incubated at 25°C for 1 h in order for the fragments to anneal, then be purified and transformed.

### 2.14 Polymerase Chain Reaction

All PCR reactions used for plasmid assembly were done using the 2x Phusion High-Fidelity PCR Master Mix with HF buffer from NEB (Wrenbeck et al., 2016) in a 50 µL volume, with the corresponding primers and sample.

In cases where the secondary structure of the sequences produced problems with the amplification, different percentages of DMSO were used (3, 6 and in one case, 9%) to help with the correct amplification. Diagnostic PCRs from colonies or phage samples were done using a 2x Taq or GreenTaq Mix from Thermo Scientific (Weyant, Edmonds, & Swaminathan, 1990) and in a 25 µl volume.

### 2.15 Goldengate assembly

(Based on Engler et al., 2008) Starting from the desired plasmid backbone, primers were designed and ordered to Integrated DNA technologies (IDT). These primers contain an overhang (a sequence not present in the original template) which introduces a BsaI/Eco31I cutting site. Compatible ends are included into the tails of the primers used to amplify the fragment to be introduced in the backbone. Using 100 fmol of each DNA template, 1 µL of BsaI from New England Biolabs (NEB), 1 µL of T4 DNA ligase from Fermentas (Rossi, Montecucco, Ciarrocchi, & Biamonti, 1997), 2 µL of 10x ligase buffer and water up to 20 µL, each reaction tube underwent the following series of cycles: 2 minutes at 37°C, 5 minutes at 16°C, (35 cycles) 5 minutes at 50°C and finally 5 minutes at 80°C.

### 2.16 Agarose gel electrophoresis

Amplified samples were run in gels made of 1% agarose in 1x TAE buffer, in order to confirm their size and approximate quantity, at 95 volts for 30 minutes as a general rule, with times varying depending on the exact size of the fragment. Smaller samples were run in a 2% gel for better discerning of their size.

### 2.17 DNA fragment purifications

All PCR reactions were purified using the ThermoFisher GeneJET PCR purification kit (ThermoFisher #K0701) according to the manufacturer's protocol. In cases where several bands appeared when running the sample in a gel, the band of interest was cut and then purified using the GeneJET Gel extraction kit from Thermo Scientific (ThermoFisher #K0691)

### 2.18 Bacterial culture

In order to obtain cell cultures for the various assays, colonies picked from grown plates were added to 5 ml of LB, with addition of the appropriate antibiotic where necessary, using Falcon 15 ml round bottom tubes. In the cases of positive selection strains, to have them ready for assays, 300  $\mu$ L of 25 mM theophylline were added to the 5 ml. Overnight cultures were refreshed the next day and used when at OD between 0.2 and 0.3. The remaining overnight culture was used to obtain glycerol stocks and minipreps for  $-80^{\circ}$  C storage.

### 2.19 Phage killing/infection

To obtain phage populations or test their killing efficiencies, phages were added to cultures within the logarithmic growth phase (OD 0.2-0.3) of the desired culture, making sure the phages would be able to properly replicate.

### 2.20 Lysate filtration

The cells post-infection were pelleted by centrifugation at 3000 rpm/2254 g for 10 min in an Eppendorf centrifuge 5920 R. To obtain clean phage samples the remaining lysates were filtered through a 2 µm Sartorius filters.

### 2.21 Bacterial Growth assays

The wells of a flat bottomed 96-well plate were filled with 180 µL of growing bacterial cell culture at an OD of between 0.2-0.3 with the addition of 20 µL of the tested phages or controls. Where theophylline was required in the culture it was added to the bacterial cells before they were aliquoted into the 96-well plate. The plate reader (TECAN SPECTRAFluor PLUS (TECAN #F121005) measured at OD<sub>600</sub> every 2 minutes after shaking for a duration of 5 hours. The growth curves for each strain were plotted using Microsoft Excel to compare differences between strains and select the best candidates for the evolution procedure.

### 2.22 Plaque assays

The efficiency of the phage replication was tested by infecting 300 µL of cells at an OD<sub>600</sub> of between 0.2-0.3 with 100 µL of phage at several different dilutions. This was then mixed with 3 ml of soft LB agar (LB agar was diluted 50% with LB) and plated with the corresponding antibiotics. After overnight growth at 37° C, clear spots known as “plaques” can be observed in the bacterial lawn of the plate. These assays were made both in the presence and absence of 1.5 mM theophylline, to assess the leakiness of the system and the activation of the riboswitch

### 2.23 One-Step assay

This procedure was adapted from Ellis, E. L. , Delbruck, (1940). These assays follow the same procedure as a plaque assay, but phage samples were obtained from infecting a culture, obtaining samples every 4 minutes over a period of 40 minutes, showing the rise in the number of phages as the infection proceeds. The assays were also made in the presence or absence of 1.5 mM theophylline, in order

to show a difference not just between the two types of sample treatment, but the addition of the molecule to the media. These phage samples were then used for plaque assays. Figure 10 shows a diagram of the procedure and sample treatment.

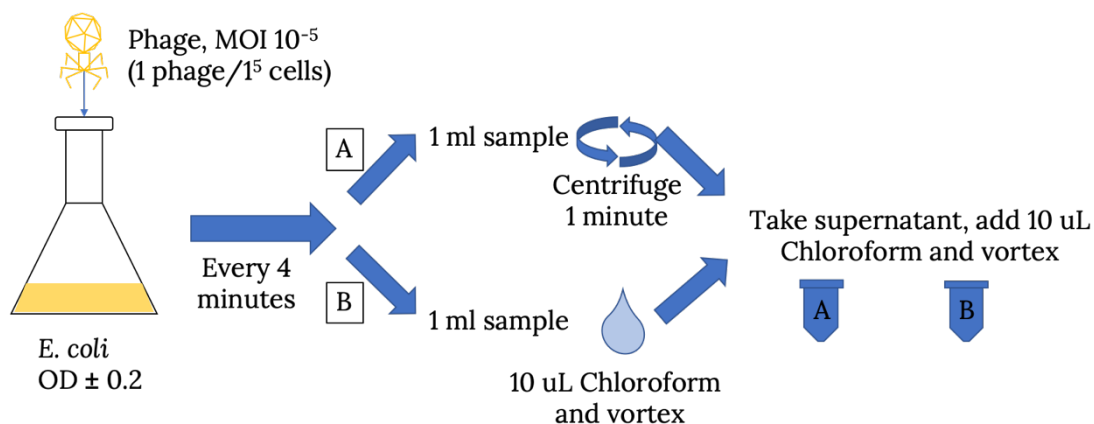


Figure 10. Overview of One-Step plaque assay.

General diagram of a one-step assay. Once the phages, at a MOI/Multiplicity of infection of  $10^{-2}$  (1 phage/100 cells) have been added to the culture, samples are taken, and undergo different treatments. “A” are centrifuged to pellet the cells and obtain the phages present in the supernatant, while “B” samples have chloroform added to them, causing bacteria to burst and release phages that had not yet exited the cells. In order to eliminate any remaining bacterial remnant present in the final samples, chloroform is added again, and those phage samples are then used for a plaque assay, where a difference should be observable between different samples and time points.

## 2.24 Next generation sequencing

In order to assess the full sizes of the riboswitch library, several different strategies and companies were considered, but finally the in-home genomics facility at the University of Warwick was chosen. A 174 bp fragment containing the riboswitch sequence was amplified via PCR from libraries in different steps of the experimental procedure, creating specific overhangs, that, using a NEXTERA XT index kit from Illumina, would be used to specifically tag the sequences of one specific library. Using an Illumina MiSeq system (Illumina #SY-410-1003), and samples from different stages of selection, the aim was to assess the library sizes in each of these steps, and then infer whether or not evolution has occurred in the process, and what that means.



## CHAPTER 3. RIBOSWITCH LIBRARY ASSEMBLY

### 3.1 Introduction

#### 3.1.1 The importance of theophylline

As explained in the introduction, riboswitches are a type of mRNA that holds a lot of promise in the field of synthetic biology, due to its structural simplicity, versatility, and regulatory capabilities. Riboswitches regulate gene expression thanks to the conformational changes produced after the binding of a specific molecule to their aptamer domain. One of the examples within this group that has been most thoroughly researched binds theophylline ( $C_7H_8N_4O_2$ ) (Chi et al., 2019; Cui et al., 2017; Nakahira, Ogawa, Asano, Oyama, & Tozawa, 2013; Rankin, Fuller, Hamor, Gabarra, & Shields, 2006; Suess, Fink, Berens, Stentz, & Hillen, 2004). This can be understood due to the fact that theophylline is a small molecule with a fairly simple structure, which makes binding easier, and that it is used for different medical treatments (Barnes, 2013). These reasons made the theophylline riboswitch an ideal candidate for a directed evolution procedure, the main objective of this project, but that could potentially be used in the directed evolution of other riboswitches and biosensor molecules.

Previous successful evolutionary experiments have been made with this riboswitch at their centre, which further support its use in this case (Lynch et al., 2007; Lynch & Gallivan, 2009; Topp & Gallivan, 2008). A library of riboswitch sequences was constructed as starting point for the evolution. There are many ways a genetic sequence library can be assembled, such as via error-prone PCR and the  $\lambda$ red system (Abou-Nader & Benedik, 2010), end-to-end ligation (Isalan, 2006) or the use of degenerate nucleotides for cassette-based PCR amplification.

#### 3.1.2 How to get away with evolution

The way this challenge was finally approached involved introducing the riboswitch sequence to be evolved inside a population of T7 phages. Due to the difficulties of directly modifying a phage's genome, the initial step was to amplify

the genetic sequences of interest with primers including a randomised region, thus assembling a library of riboswitches. This fragment was then cloned in a plasmid backbone and transformed into *E. coli* cells, which were then infected by WT T7 phages, some of which incorporated the sequence via the process known as homologous recombination (Haber, 2008).

Gallivan and collaborators had already tested a similar approach for the evolution of a random riboswitch sequence (Lynch et al., 2007), creating randomised libraries of the theophylline riboswitch with different spacer sizes between its end and the beginning of the regulated gene, as well as a high-throughput motility-based screening to detect optimal riboswitches. This research, along with the previously mentioned PACE strategy, are essential influences in the development of this project. Previous work (Lynch & Gallivan, 2009) showed the possibility of doing the same with a 12 nucleotide (12N) fragment, but given how this project was focused on developing the method, such a wide variety of sequences was deemed excessive, and the decision was taken to just randomise a region of 8 nucleotides.

To help proper regulation, upstream from the riboswitch regulatory elements such as a T7 terminator and promoter (Temme, Hill, Segall-Shapiro, Moser, & Voigt, 2012) were included, to control the expression of the  $\lambda$  phage *cI* gene (Fehér et al., 2012), which will be essential in the selection process the riboswitch sequences would have to go through.

Although the project was initially designed with homologous recombination in mind, over the course of time an alternative was also considered, via the inclusion of a T7 packaging signal. Yosef, Goren, Globus, Molshanski-Mor, & Qimron, (2017) showed that by transforming a cell with a plasmid carrying a phage's packaging signal, WT phages infecting those cells were able to pack the plasmid inside the virion instead of their original genome. This process creates what is known as a phage particle or transducing particle. Their research demonstrated changes in host specificity by introducing genes corresponding to phage-host interaction, showing that foreign genes can function appropriately in this type of particle.

Following their steps, and previous research about T7 packaging signal such as Chung & Hinkle, (1990), the approach for the original phage library of riboswitches was reconsidered. Instead of undergoing homologous recombination to introduce the riboswitch library inside the phage, homology regions would be eliminated from the plasmid, and a T7 packaging signal would be included along with the library and the control riboswitch.

By doing so, the theory was that it would allow for some advantages when compared to the full phage system, such as a higher efficiency than with homologous recombination, in itself an uncommon occurrence. But most importantly, it gave the whole system, specifically the riboswitch, more time in the presence of theophylline as the cells were growing.

This would translate into higher activation levels of *cI* expression by the time of infection, instead of the more or less 17 minutes it had when present inside of the phage. In the latter, the whole process of interaction with theophylline and *cI* regulation had to occur between phages attaching themselves to the cell and bursting out at the end of the lytic cycle, a process which showed useful differences in activation levels, but that could be improved.

Apart from the introduction of a packaging signal, this new plasmid library required an alternative regulation system (since the phage will only be acting as a vessel); meaning that the original T7 promoter placed in front of the riboswitch had to be exchanged for an *E. coli* promoter. Due to its already known strength and efficiency, the chosen part was J23119 (C. C. Liu et al., 2012).

### 3.2 Specific Objectives

To obtain an initial plasmid library of randomised riboswitches. The plasmid library could then be turned into a phage library via homologous recombination or have a T7 packaging signal incorporated into its sequence, to allow packaging inside the T7 virion transduced into other cells.

### 3.3 Specific materials and methods

#### 3.3.1 Homologous recombination/Full phage procedure:

- Vector assembly:

Using as a template the plasmid EG\_001, which contained regions of homology with the T7 genome flanking the  $\lambda$  cI gene and the essential genes *trxA* and *cmk*, several of the primers from Table 5 were designed to include, via PCR, a gene fragment containing a T7 terminator, T7 promoter, and a truncated riboswitch sequence, all incorporated in front of cI, ensuring its expression was regulated by the activation of the working riboswitch. Before all this was attempted, *cmk* was removed from the plasmid using the compatible restriction enzymes BglII and HindIII, as its presence within the phage would render the designed selection process useless.

Once *cmk* had been eliminated from the plasmid, the assembly of the final desired plasmid containing all the elements needed required several intermediate steps. Plasmid EG\_001 was amplified via PCR using primers EG\_GG001 and EG\_GG002, which stand out from other primers due to their significant length. This was done so as to make the cloning procedure easier, allowing for the introduction of several elements in one step of cloning.

This amplification created intermediate sequence EG\_temp01 (Supplementary figure 1A). Furthermore, both primers were designed with cutting sites for BpiI/BbsI at the end of their sequence, as shown in Figure 11, so that after amplification, both ends of the plasmid could be digested and re-ligated via Gibson assembly, as described in the general materials and methods. After this procedure, the re-ligated plasmid EG\_temp02 was obtained (Supplementary figure 1B). This vector contained all the necessary elements for the experiments, with the exception of the specific randomised region within the riboswitch that will be the base of the library, as well as the fixed version of the riboswitch that was to be used as a control. In order to obtain this, the vector had to be cut using the incorporated Eco31I/BsaI sites introduced through PCR, and then set up for GoldenGate reaction with the desired fragments, as previously described.

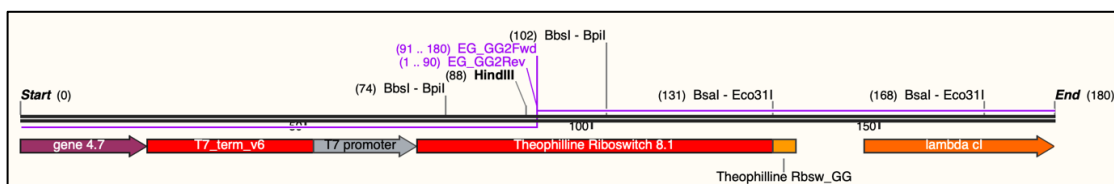


Figure 11. Initial amplification for HR fragment.

Detail from the sequence added via the use of primers EG\_GG02Fw and EG\_GG02Rev, which also incorporate cutting sites for digestion enzymes BpI/BbsI and BsaI/Eco31I for further digestions down the working pipeline.

- Primer Overlap:

In parallel to these procedures, two nucleotide fragments were assembled, using three overlapping oligos (EG\_GG03, EG\_GG04 and EG\_GG05) which also introduced within the sequence the Eco31I/BsaI sites complementary to the ones included within the plasmid. One of the sequences contained the ending nucleotides of the 8.1 version of the theophylline riboswitch, from (Lynch et al., 2007) which had already been shown to be a very effective version of the riboswitch, with a high activation rate (EG\_GG04 and EG\_GG05). The other fragment was made using degenerate oligos, containing eight random nucleotides (8N) in an even distribution of 25% for each nucleotide, giving a final sequence space of  $65.536 \times 10^4$  different sequences to begin the evolutionary process with (EG\_GG03 and EG\_GG05). To produce these fragments via PCR, for the Golden Gate assembly, each of the two different combinations were mixed along Phusion High-fidelity PCR MasterMix and went through a specific protocol on the Thermocycler, with 5 rounds of 30 seconds at 70°C + 10 seconds of extension at 72°C. These reactions were then purified using the previously described PCR purification kit and eluted in 10 µl. Both fragments are shown on Figure 12.

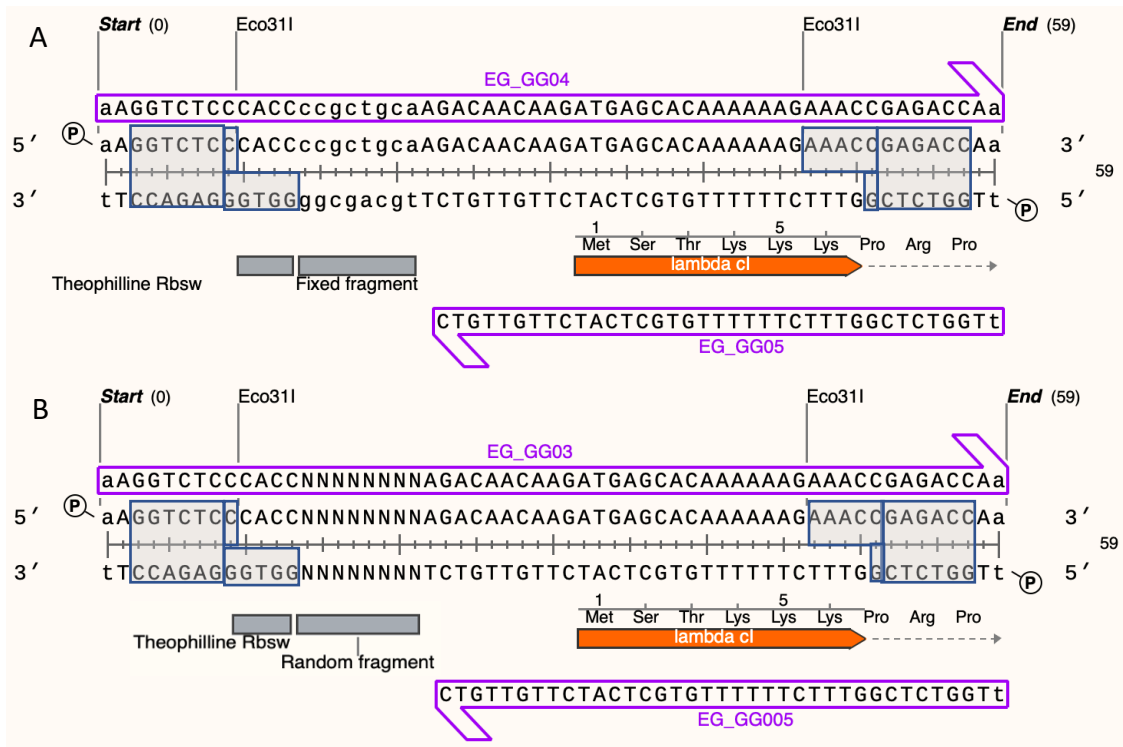


Figure 12. Fixed and random riboswitch fragments inserted in HR fragment.

Details from the fragments produced via oligonucleotide annealing of the primer pairs EG\_GG004 and EG\_GG005, and EG\_GG003 and EG\_GG\_005. Both have the same length and cutting sites, with the main difference being that the sequence of 04 (A) includes a fixed tail for the theophylline riboswitch, while the one from 03 (B) contains 8 random nucleotides that form the random riboswitch library. The grey squares represent the cutting sites for the enzyme Eco31I/BsaI, designed to be complementary to the ones present in the plasmid where the fragments will be incorporated.

These two sets of fragments were then incorporated into the plasmid backbone of EG\_temp02 via GoldenGate assembly, using the complementary BsaI/Eco31I cutting sites present in the plasmid and the fragments to introduce the final nucleotides of the riboswitch within the plasmid without leaving a digestion scar.

- Homologous recombination:

The production of a plasmid library, however, is not the main objective of this project, since to evolve the riboswitches, they need to be inserted into the genome of a T7 phage. In order to do so, these two plasmid populations were each transformed into *E. coli* BW25113 cells which already contained the T7 gene

product 5 (gp5: DNA polymerase) in a plasmid backbone carrying Kanamycin (Kan) resistance, thus making these cells Kan and Cm resistant.

Once these populations were established, they were subjected to homologous recombination (Haber, 2008), being infected with WT phages so that during the replication procedure, the regions present in the plasmid which are homologous to the phage's genome allow for the exchange of a fragment between phage and plasmid. This exchange produces a subpopulation of phages different from WT phages, which will be carrying the fragment previously flanked in the plasmid by the phage's homology arms.

In order to maximize this procedure, there must be enough phages so as to allow proper replication and avoid bias in the phage population, while at the same time not overwhelming bacteria by completely killing them in one round of replication; allowing for several rounds to augment the chances of the recombination process. After several tests with different ratios, we arrived at an ideal phage/cell ratio of  $10^{-5}$ . This can be seen in Figure 16 in the results section.

Due to the loss of the T7 DNAPol in their genome, in order for these phages to be obtained, it was necessary for the cells to carry a copy of the previously mentioned T7 gp5/DNA pol gene. The presence of this protein is essential to ensure the replication of any phage that undergoes homologous recombination, as they lose their WT copy of gp5 and thus their ability to replicate when incorporating either version of the fragment within their genome.

- Selection of homologous recombined phages:

The resulting phage lysates obtained from these infections were comprised of both WT and homologous recombined phages, so in order to obtain a pure stock, they needed to be selected. Selection was achieved by infecting *E. coli* BW25113  $\Delta trxA$  T7gp5+ cells with the lysates, as WT phages would not be able to replicate due to the lack of the thioredoxin A gene, while recombined ones carry it as part of the fragment of interest. These initial lysates were used for plaque assays in both  $\Delta trxA$  and regular *E. coli* BW25113 cells, in order to assess the presence of



WT and recombinant phages. The plaques formed in the first case were isolated and then used to infect liquid cultures of  $\Delta trxA$  cells. The lysates obtained from these infections were formed exclusively of recombined phages, and this hypothesis was once again tested via plaque assays in both  $\Delta trxA$  and regular *E. coli* BW25113 cells, confirming the absence of WT.

- Riboswitch sequencing:

Sanger sequencing provided by Eurofins Genomics was used at different points throughout this process to ensure the presence of the desired riboswitch variants. Each sample sent contained on average 400 ng of DNA in a total volume of 10  $\mu$ l, including 1  $\mu$ l of the specific primer used. Sequencing was done in batches after Gibson assembly of the plasmid populations, and after the homologous recombination procedures. While multiple bacterial clones and phages were individually sequenced and several different nucleotide fragments were found within the random region, it needs to be stated for the record that the full scale of the variation present in the random riboswitch library was not assessed at the initial point of the experimental pipeline. The full scale of variation was determined much further down the road, as the proper setup and development of the evolution process was prioritised, given that only the Fixed control riboswitch sequence was needed in order to properly test the methodology.

### 3.3.2 Transduction/Phagemid strategy:

- Packaging signal design:

Based on the work by Yosef, Goren, Globus, Molshanski-Mor, & Qimron, (2017b) and the original T7 packaging signal reference used by them (Chung & Hinkle, 1990b, 1990a), different packaging signals were designed with the purpose of integrating the sequence inside of the riboswitch plasmids EG\_cI01 and EG\_cI02. These sequences were formed of several T7 elements and ordered from IDT DNA in the form of readily assembled Gblocks. Such fragments included a T7 promoter PhiOR (Orlov, Ryasik, & Sorokin, 2018), short repeats from the phage's genome right end (SRR) (Studier & Dunn, 1982), one set of the phage's terminal repeats (TR) (Hoelz, Hickey, & Malkas, 2004) and the short terminal repeats from the left side (SRL). The different sequences simply varied in the length of these



different elements, with the differences being shown in Figure 13. Only one of the two identical copies of the terminal repeat sequence is included within the packaging signal so that it would hybridise with the ones present in the phage's genome, in a process of concatemer formation design to avoid DNA loss during replication.

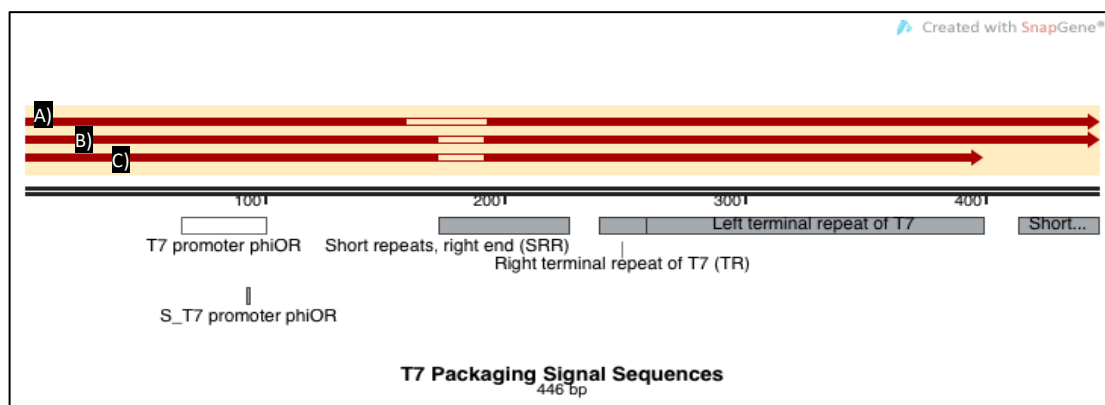


Figure 13. Structure of the different packaging signals.

Different T7 packaging signal sequences designed over the course of the project. In grey, each of the elements incorporated as part of the different packaging signals, acting as a reference. In red, the parts of each sequence that aligned to the genetic sequence of each reference. The middle gap in each of the three sequences represent a bses present in the standard that were not incorporated into the novel sequences. Sequences B and C were the initial attempts, differing in the inclusion of the left end short repeats (SLR), while sequence A was the one that was finally incorporated into the riboswitch plasmid.

- Vector assembly:

Starting from the plasmid EG\_cI02 and using the process of GoldenGate assembly, sequence A from Figure 13 was incorporated into the plasmid, eliminating all but the last 29 nucleotides of T7 homology region 4.7. Using the same technique, gene 5.3 was eliminated, introducing in its place a copy of the error-prone T7 DNA polymerase, which would allow, as stated in the introduction, for the  $\Delta$ gp5 phages used in this strategy to replicate inside the cell and avoid having to use WT T7 in the process. The plasmid also retained the original Cm resistance.

No modification of EG\_cI01 was made, since by this point the ideal selection strains had already been identified and would be used without any further

modification in this process. Because of this, there was no need for a control plasmid carrying the fixed version of the riboswitch in this strategy.

This gave as a result the sequence known as EG\_Pm01b (Supplementary figure 8. A). This plasmid was tested for some phagemid production experiments but was discarded and some of its elements replaced due to a basic flaw that had gone unnoticed. The remainder of gene 4.7 was eliminated, and the T7 promoter included in front of the riboswitch was substituted for the *E. coli* promoter J23119, giving as a result EG\_Pm01C (Supplementary figure 8. B)

- Phagemid library production experiments:

In the initial experiments, to test the procedure, cultures of cells carrying the packaging signal plasmid were infected with WT phages until the culture cleared out, after approximately 1 h. Then, the tubes were centrifuged at 14000 *g* for 10 minutes, and supernatants were filter purified. These supernatants were treated as phage lysates, and assays were developed as classical plaque assays, save for the absence of soft agar. 300  $\mu$ l of cells at an OD between 0.2 and 0.3 were mixed with 100  $\mu$ l of the lysate and then plated onto Cm-containing plates, which were left to grow overnight at 37° C. If phagemids were present in the lysate and managed to infect the cells, colonies would appear on the plate. As a parallel assay, in order to assess the presence of WT phages, regular plaque assays were done, as previously described. The efficiency of the process was measured by the ratio of colonies to plaques in the 2 assays. These assays were eventually carried out with multiple types of cell strains and phages, including *E. coli* strains TOP10, BW25113, BW25113  $\Delta$ *trxA*, BW251143  $\Delta$ *cmk*, and BW25113 $\Delta$ *cmk* $\Delta$ *trxA*. The phage strains include WT T7 phages, and two types of  $\Delta$ *gfp5* T7, one carrying *trxA* and *cl*, and one carrying *trxA*. All these combinations were assessed to confirm the exact conditions where the transduction process could work, and whether or not the absence of the essential genes would impair the function of a phage particle as it did in the case of HR phages. TOP10 were used as a negative control of sorts, as previous lab experience shows T7 infection is less efficient in this particular strain than it is on BW25113.

### 3.4 Results:

### 3.4.1 Plasmid assembly

- HR\_Fix plasmid/EG\_cI01: 3569 bp (Figure 14A)

This plasmid incorporated the same final nucleotides present in the 8.1 version of the theophylline riboswitch shown in Figure 14B, (from Lynch et al., 2007). This plasmid served as the quality control for the system, carrying a properly established riboswitch, and using it to set up the different steps before attempting evolution by using the randomised library of switches.

- HR\_Random plasmid/EG\_cI02: 3569 bp.

Comprised of the same elements as EG\_cI01 (Figure 14A) , but in this construct the riboswitch contains a randomized region of 8 nucleotides after the stem of the riboswitch and before the RBS (Figure 14C). This means that this is actually a plasmid library of riboswitches, containing theoretically  $65.536 \times 10^4$  variants. These will be used for the evolution experiments once the method has been properly set up using the fixed version of the riboswitch.

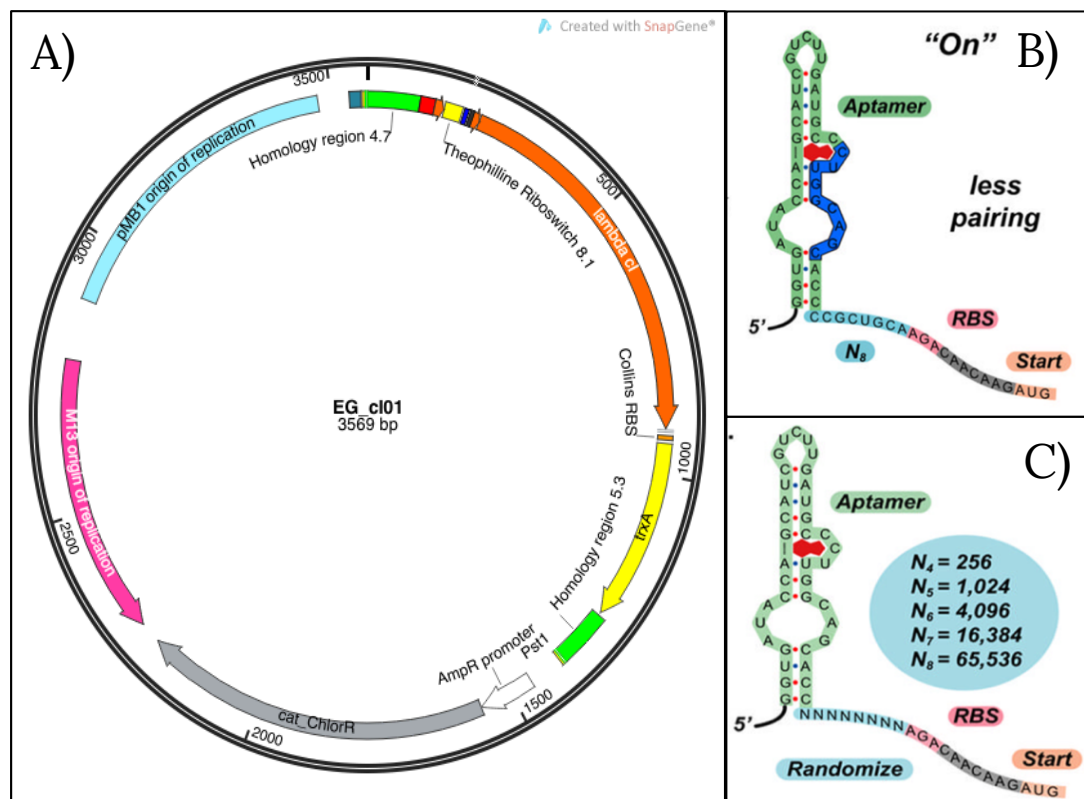


Figure 14. Homologous recombination plasmid carrying riboswitch variants.

Riboswitch-carrying plasmid used for homologous recombination. A) gene map indicating elements carried by the plasmid, with the homologous recombination fragment present between gene 4.7 and gene 5.3. B) 8.1 version of the riboswitch, which showed the biggest activation fold in Lynch, et al., 2007. C) Structure of the theoretical library obtained after randomizing the 8 nucleotides following the stem of the riboswitch.

Both plasmids had the exact same size and components, and only differ in those 8 nucleotides at the end of the riboswitch. Details of the final riboswitch sequence between the T7 terminator and the  $\lambda$  cI gene are shown in Figure 15.

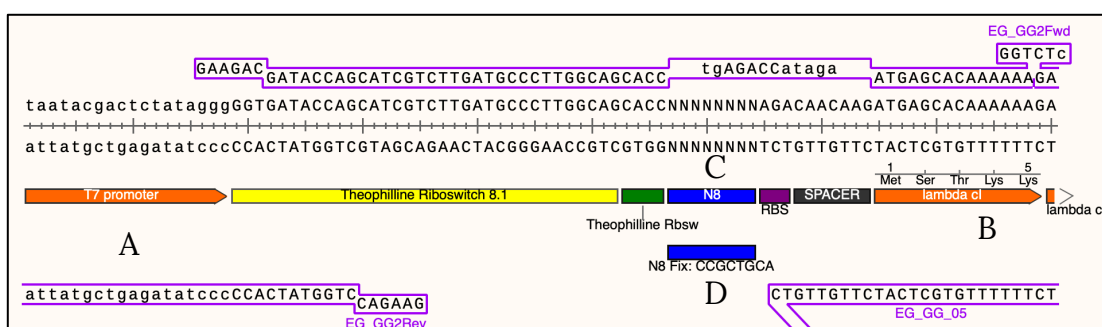


Figure 15 Final homologous recombination fragment

Final sequence of the modified region between the T7 promoter (A) and the cI gene (B), showing the two final versions of the riboswitch, the Fixed one (C) and the Randomized one (D). It is also shown that the pre-existing Eco31R/BsaI sites are gone after the GoldenGate reaction took place, which explains why part of the primers used to incorporate them (EG\_GG2Fwd & Rev) do not bind to the sequence anymore. The full sequence of these primers can be found in Table 5.

### 3.4.2 Homologous recombination procedure:

A PCR of the same region in the phages obtained at different MOIs showed the difference in size between fragments from WT, erroneous HR phages, and correctly recombined phages, easily discernible in Figure 16.

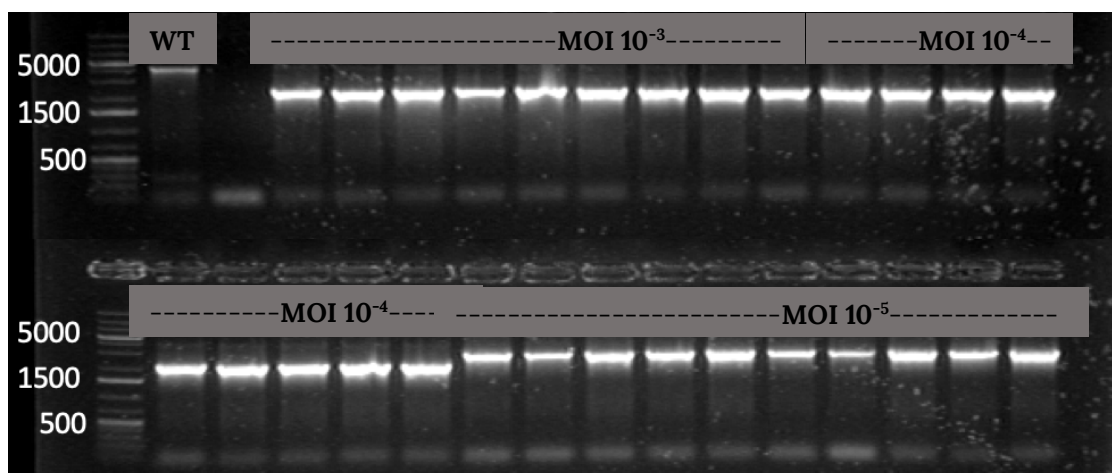


Figure 16. Results from HR procedure at different MOIs.

The bands correspond to the amplification of the homology region between T7 gene's 4.7 and 5.3. The image shows the size for the original WT amplification of T7 *gp5* (4058 bp) as a control, and the results from the HR procedure at different MOIs. Both  $10^{-3}$  and  $10^{-4}$  show a band at roughly 2000 bp, smaller than the expected size, indicating something went wrong in the procedure, while  $10^{-5}$  shows a fragment of the expected size (3017 bp), corresponding to the full fragment of interest. The ladder used was the GeneRuler 1kb Plus ladder.

### 3.4.C Library assembly

Samples from the different plasmid constructs were sequenced to confirm whether or not they contained the correct sequences. After the homologous recombination process, 20 phages that appeared to have undergone the process were also sequenced to assess if they indeed contained the desired sequence of interest. The results from the sequencing of elements from EG\_cI01 and EG\_cI02, as well as these recombinant pre-selection phage populations, show the different versions of the riboswitch in both the plasmid and phage libraries. The Fixed, non-randomised version of the switch was present in the phage and plasmid populations, confirming it remained unchanged. Both the plasmid and the phage populations containing the randomised libraries show no repetition of sequences, which gives us basis to believe there is no bias in the population, and all 8N sequences are similarly represented. All of these sequences can be seen in Figure 17.

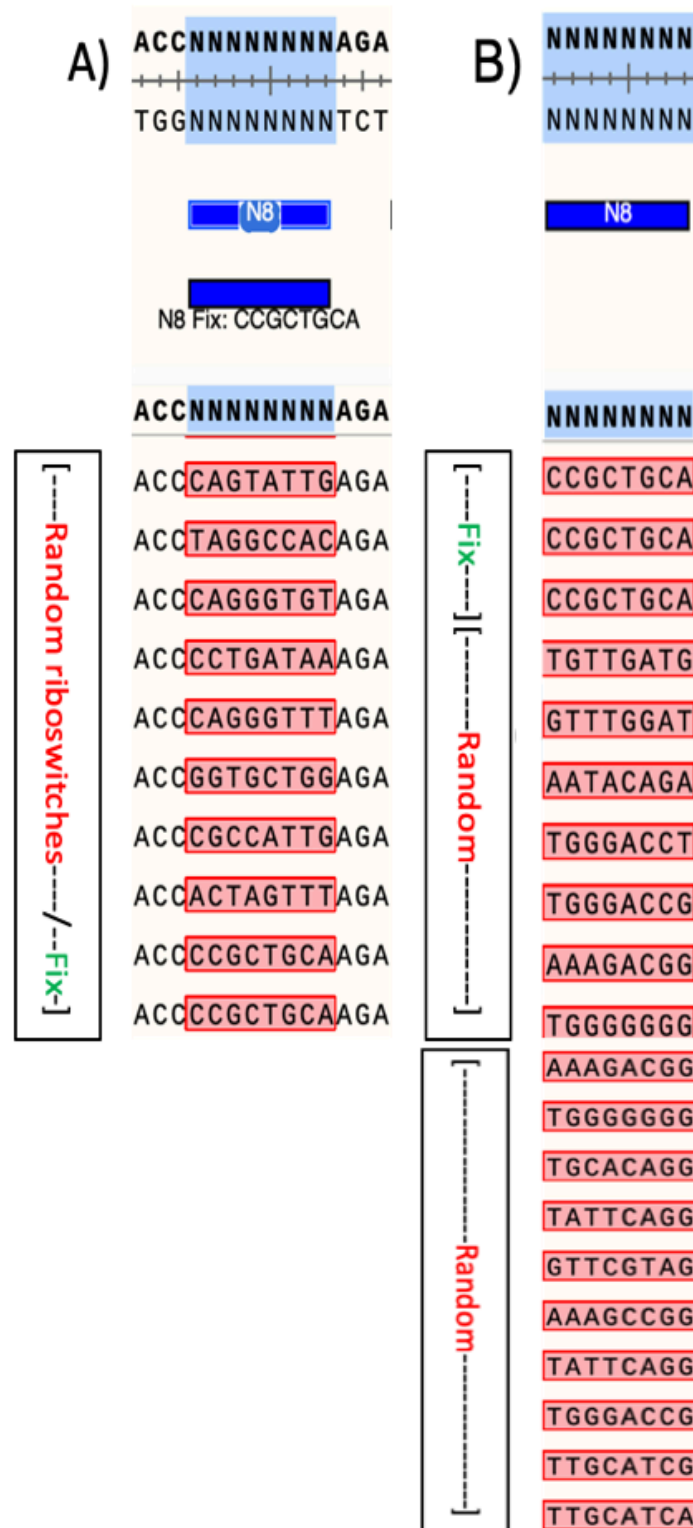


Figure 17. Sequence variation within obtained plasmid and phage populations.

Collection of sequences obtained throughout the development of the different libraries.

A) Riboswitch sequences obtained by Sanger sequencing corresponding to the PCR of single colonies carrying plasmids EG\_cI01 (Fix) and EG\_cI02 (Random).

B) Sequences from the PCR of HR phages carrying the fragment of interest, which could come from EG\_cI01 or EG\_cI02. Fix phages were used to set up the evolution procedure.

The results from the sequencing of EG\_cI01 and EG\_cI02 are also applicable for the alternative method of transduction, as the plasmid library had already been established, and the only difference was the addition of the different packaging signal sequences to test.

### **3.5 Discussion of results:**

The obtained sequencing results from both plasmid populations confirmed the presence of the sequence of interest, both in the case of the fixed riboswitch sequence, where it remained constant in all members, and the randomised riboswitch library, with no sequence appearing more than once in the different samples analysed.

The subsequent analysis of the recombined phages proved that the populations of phages had incorporated the sequences in both cases, and the variation present in the case of the randomised phage library indicated no bias. This allowed the project to go on to its next stage and use the phages carrying the fixed riboswitch to set up the evolutionary process, and then test the random phage to evolve the sequence over generations of selection.

Further down the experimental timeline, Next Generation Sequencing procedures took place, and samples corresponding to EG\_cI01, cI02 and the initial phage populations were prepared to be run in an Illumina MiSeq and sequenced, in order to further confirm the library variation within each population. These results can be seen in Chapter 4 of this thesis.



## CHAPTER 4. RECOMBINATION-BASED EVOLUTION STRATEGY

### 4.1 Introduction:

#### 4.1.1 T7, the perfect tool for evolution.

The T7 phage is a member of the T4 family of bacteriophages, and one of the most researched and understood phages. This is mainly due to its obligatory fast-paced lytic life cycle, which covers cell entry, replication, assembly and lysis in around 17 minutes, generating around 100 offspring per infection (Molineux, 2005). Its fully mapped genome is comprised of 40 kb, and the function all the key genes are appropriately accounted for (L. Y. Chan et al., 2005).

A schematic of the genome of the T7 phage and its virion are shown in Figure 18A and B. The full list of the genes in the T7 genome can be found in in the Supplementary Materials and Methods section.

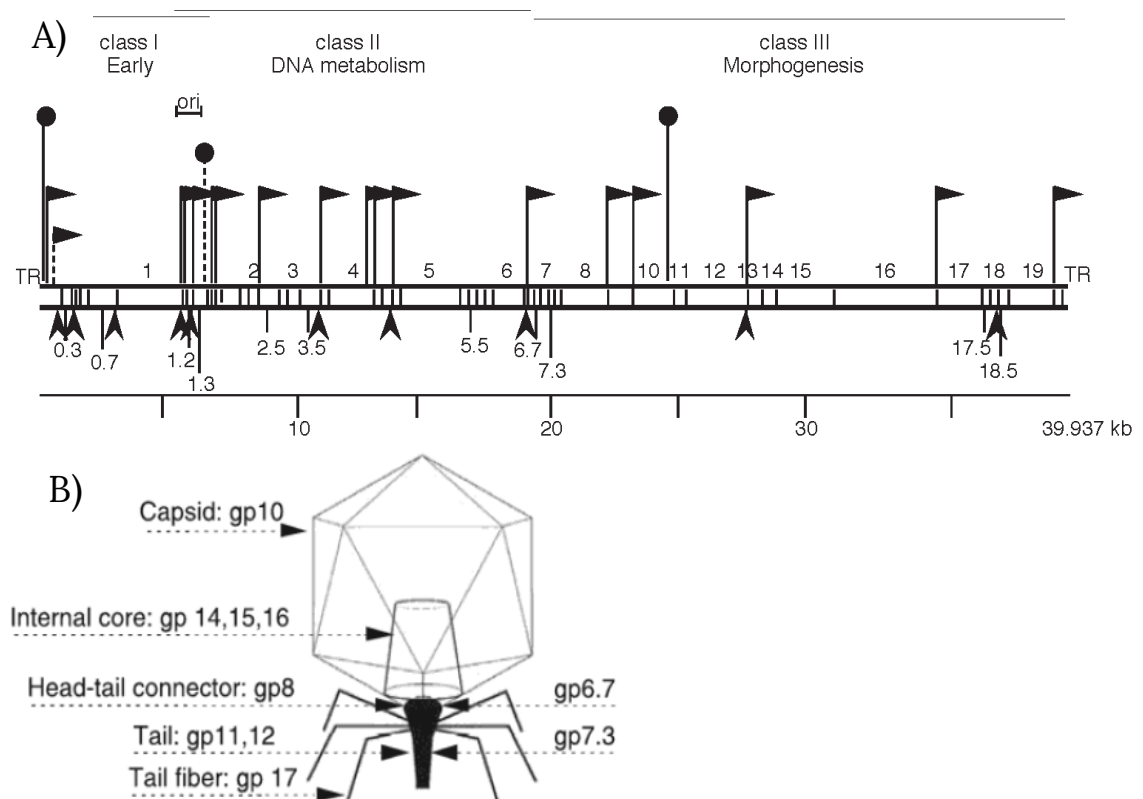


Figure 18. Genome and virion structure of T7 phage.



A) Schematic diagram for the genetic structure of the T7 phage. The genes are progressively transcribed over time, starting from the left of the diagram to the right. Promoters are marked with a black arrowhead, while terminators are marked with a black circle. Arrowheads below the map represent RNase III processing sites, and the short lines between the long, parallel ones represent the beginning and end of each gene, marked above by its corresponding number. B) Schematic representation of the T7 phage's virion, representing its protein components, encoded by the genes marked as "Class III" in A). Both figures were taken from (Molineux, 2005).

An advantage of studying the T7 phage is that its mutation rate can be regulated through the replacement of the gene encoding its DNA polymerase with an error-prone version (Friedberg et al., 2001). This makes it ideal for a directed evolution strategy, such as the one discussed within this thesis.

### 4.1.2 Riboswitch evolutionary strategy: EvolutiON & EvolutiOFF

As discussed previously, there were two strategies developed during the duration of this project. One, based on the use of T7 phages, and another based on T7-derived infected particles, or phagemids.

The main strategy, designed and tested during the project, was based on full T7 phages and homologous recombination, to achieve the evolution of a riboswitch sequence by tying it to the phage's replication efficiency. After homologous recombination, the resulting phages lose their gp5/DNA polymerase in favour of the fragment of interest (Figure 15 in Chapter 3). gp5 is essential for the phage to replicate and therefore the loss of this gene from the phage genome is a problem. This is overcome by expressing the gp5 gene *in trans* inside the infected cells.

In the fragment of interest, a T7 promoter and the riboswitch regulate the expression of  $\lambda$  phage gene *cI*; upstream from the gene *trxA*, which codes for the protein thioredoxin A. this is an *E. coli*-produced protein essential for the phage's replication cycle, which works as the processivity factor for the phage's DNA polymerase (Qimron et al., 2006). This gene, however, is not the main focus of the evolutionary procedures, this corresponds to the riboswitch.

Riboswitches have two 3D configurations dependant on the presence of their specific activator molecule, which in this case is theophylline. Therefore, selection needed to have two steps: 1. Selection is designed towards an “ON” stage of the switch, with the activator present, and 2. Selection towards an “OFF” state, with the activator absent. Each of these selections occur in *E. coli* selection strains with their own genetic makeup. In both selection steps, the cells were KO for the *trxA* gene, as it was already included within the T7 phage. In the positive selection, the cells also lacked *cmk*, another T7 phage essential gene.

Genes *trxA* and *cmk* were selected as targets in this strategy due to their large impact in the proper replication of the T7 phage. A full list of *E.coli* genes with an impact on T7 replication (from Qimron *et al.*, 2006) can be seen in .

Table 6. List of *E. coli* genes identified as essential for T7 growth.

From Qimron, et al, 2006. Table shows the names, gene products and T7's efficiency of plating (EOP, essentially the titre of the phage preparation); both when the gene is knocked out and when it is complemented via a plasmid.

Gene	Gene product	EOP	EOP after complementation
<i>trxA</i>	Thioredoxin A	<10 <sup>-2</sup>	1
<i>cmk</i>	CMK/dCMP kinase	0.05	0.91
<i>gmhA</i>	LPS biosynthesis enzyme	<10 <sup>-2</sup>	0.86
<i>gmhB</i>	LPS biosynthesis enzyme	0.30	0.83
<i>waaC</i>	LPS biosynthesis enzyme	<10 <sup>-2</sup>	0.80
<i>gmhD</i>	LPS biosynthesis enzyme	<10 <sup>-2</sup>	0.16
<i>gmhE</i>	LPS biosynthesis enzyme	0.01	0.88
<i>waaF</i>	LPS biosynthesis enzyme	0.23	0.79
<i>waag</i>	LPS biosynthesis enzyme	0.01	0.82
<i>galU</i>	LPS biosynthesis enzyme	0.29	0.82
<i>waaR</i>	LPS biosynthesis enzyme	<10 <sup>-2</sup>	0.79

The *cmk* gene encodes CMP/dCMP kinase, which catalyses the conversion of CMP/dCMP to CDP/dCDP, using ATP as a phosphate group donor. The lack of constitutive *cmk* in the cells is explained by the fact that its expression is what controls the positive selection step. When a phage containing a random riboswitch sequence infected a bacteria from the positive strain, if the riboswitch

attained its “ON” conformation, the transcription of the *cl* gene would start. The repressor would then bind to the pRM promoter in the cell and activate the expression of *cmk*, allowing phages with a working version of the random riboswitch to replicate. A schematic of this process is shown in Figure 19.

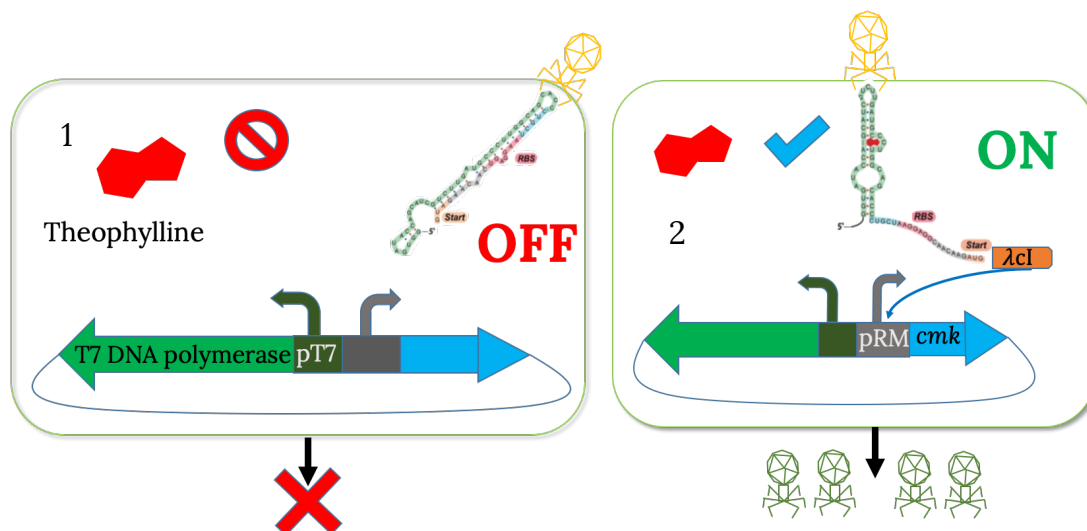


Figure 19. Schematic representation of the positive selection process.

As the aim of this step is to select riboswitches in an “ON” state, theophylline is present in the media where cells are grown. 1) If the phage carries a version of the switch which is unable to turn on in such a case, the cascade initiated by *cl* won’t start, and no *cmk* will be produced, thus impairing phage replication and eliminating those members from the population. 2) If the switch is able to turn on in this case, however, phage replication will occur and those elements will remain in the population.

All phages able to replicate in the positive selection cells would go on to infect the negative selection strain. In this case, with the activator molecule absent, the objective is to select for riboswitches that remain in an “OFF” conformation. In this case, the pRM promoter regulated the expression of the T7-exclusion gene *pifA*. This gene was originally found on the *E. coli* F-plasmid (Garcia & Molineux, 1995) and is able to block the replication of T7 when it is expressed. The mechanism by which this process occurs is not well defined, but it is known that it is triggered by the co-expression of *pifA* and T7 genes 1.2 or 10 (Schmitt & Molineux, 1991), and that it happens on the surface of the cell (Cheng, Wang, & Molineux, 2004). By conditionally regulating the expression of this gene, any phage containing a version of the riboswitch with a constitutively “ON”

conformation would be eliminated from the system; and only those which are “OFF” in the absence of theophylline would be able to replicate and stay in subsequent evolution steps. A schematic of this process is shown in Figure 20.

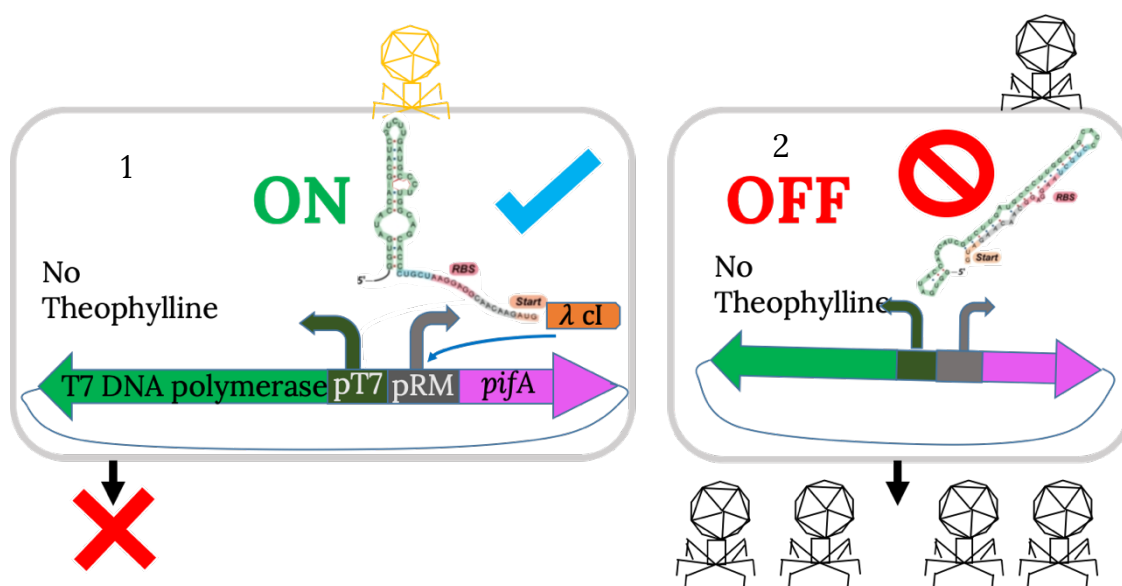


Figure 20. Schematic representation of the negative selection process.

As the aim of this step is to select riboswitches in an “OFF” state, theophylline is absent in the media where cells are grown. 1) If the phage carries a version of the switch which remains on in such a case, the cascade initiated by *cl* will start, and *pifA* will be produced, thus impairing phage replication and eliminating those members from the population. 2) If the switch remains off in this case, however, phage replication will occur and those elements will remain in the population.

The presented experimental setup was also valid for the evolution of other types of sensors, with the main elements driving the selection (*cmk*, *pifA*) being the same in all cases. The *mmfR*/*mmyB* genes are part of a larger system that regulates biosynthesis of the antibiotic molecule known as methylenomycin (Mm) in *S. coelicolor*, and are autoregulated by the presence of molecules known as methylenomycin furans (MMF), also produced within the system (O’Rourke et al., 2009). The *mmyB* gene codes for a transcriptional activator of the synthesis of Mm, and in order to regulate the system, MmfR represses *mmyB*. However, once concentration of MMF molecules is high enough, MmfR repression is blocked.

A simplified diagram of the pathway can be found in Figure 21.

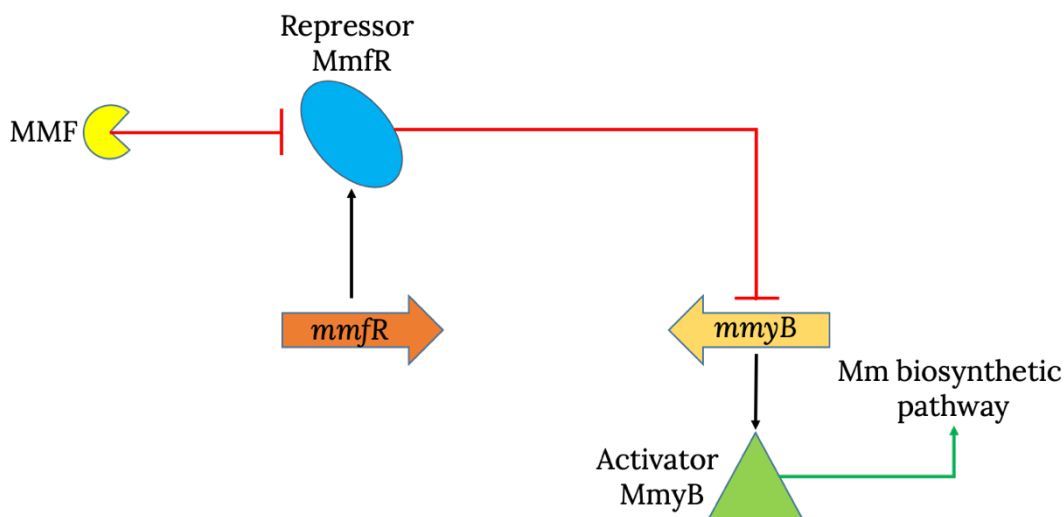


Figure 21. Diagram of the interactions between MMF, *mmfR* and *mmyB*.

Genes *mmfR* and *mmyB* are part of the regulating system responsible for Mm production. *mmfR* encodes MmfR, which represses the *mmyB* and thusly, the biosynthesis of the Mm compounds. Once the levels of MMF are high enough to repress MmfR regulation, the MmyB activator can be produced and Mm biosynthesis starts.

To this end, T7 phages underwent homologous recombination and selection, in order to obtain phages containing the 3 different versions of the *mmfR* (transcribed from original gene *mmfr* (O'Rourke et al., 2009), and mutants SAV2270 (Corre, Song, O'Rourke, Chater, & Challis, 2008) and *sgnR* (Sidda et al., 2014). These would be used in the selection process, with a plasmid containing the MARE4 gene variant and *cmk* in the positive selection step, where the repressor would not bind and thus the gene will be expressed; and *mmyB* and *pifA* in the negative selection, where the objective would be to bind the repressor and block the production of the exclusion protein PifA.

All information regarding these processes and the results obtained will be discussed in the following pages.

## 4.2 Specific Objectives

The objective of this chapter was the development and further assessment of a methodology for the directed evolution of riboswitches using recombined T7 phages, and to adapt this methodology for the evolution of alternative sensors.

## 4.3 Specific materials and methods:

### 4.3.1 Plasmid assembly specifications:

All the constructs assembled were based on previously available plasmids present in the laboratory's stock. The assemblies were made either by enzymatic digestion at 37°C and overnight ligation at 20°C, or using GoldenGate ligation (Engler et al., 2008), where indicated as such.

- EG\_001: 3490 bp (Figure 6)

Starting from the backbone of this plasmid, PCR amplifications were done to incorporate GoldenGate cutting sites, which would allow for a gene fragment to be incorporated, containing a T7 promoter sequence controlling the riboswitch, the theophylline riboswitch, and the *ci* gene. This process was done for both EG\_001 and EG\_002, in order to have both the control and library populations.

- EG\_002+, EG\_002-: 7135/8677 bp, respectively (Figure 8 and Figure 9)

These original plasmids underwent further modifications in order to assemble all the positive and negative selection plasmids. All different cargos were introduced into EG002 via Goldengate assembly, using the previously mentioned GoldenGate cutting sites.

Plasmids specific to the  $\lambda$  *ci* project:

- HR\_Fix plasmid/EG\_001: 3569 bp. (Figure 14)

As described in the Library Assembly chapter, this plasmid was based upon the  $\Delta$ *cmk* version of Pr100 pLit Chlor previously described in general methods. Via the use of GoldenGate assembly, a gene fragment containing a T7 terminator, a T7 promoter, and a version of the theophylline riboswitch with a high-fold

activation was introduced directly upstream from the *ci* regulator from phage  $\lambda$ . This riboswitch was taken from the literature (Figure 22A, from Lynch, Desai, Sajja, & Gallivan, 2007). This plasmid was used as a positive control for the efficiency of the method; given that since it carries a properly established riboswitch, which we knew turned ON and stayed OFF in the corresponding environments.

- HR\_Random plasmid/EG\_cI02: 3569 bp.

The same elements as EG\_cI01, but the riboswitch contains a randomized region of 8 nucleotides after the stem of the riboswitch and before the RBS (Figure 22B) creating a plasmid library of riboswitches containing theoretically  $65.536 \times 10^4$  variants. These will be used for the evolution experiments once the method has been properly set up using the fixed version of the riboswitch.

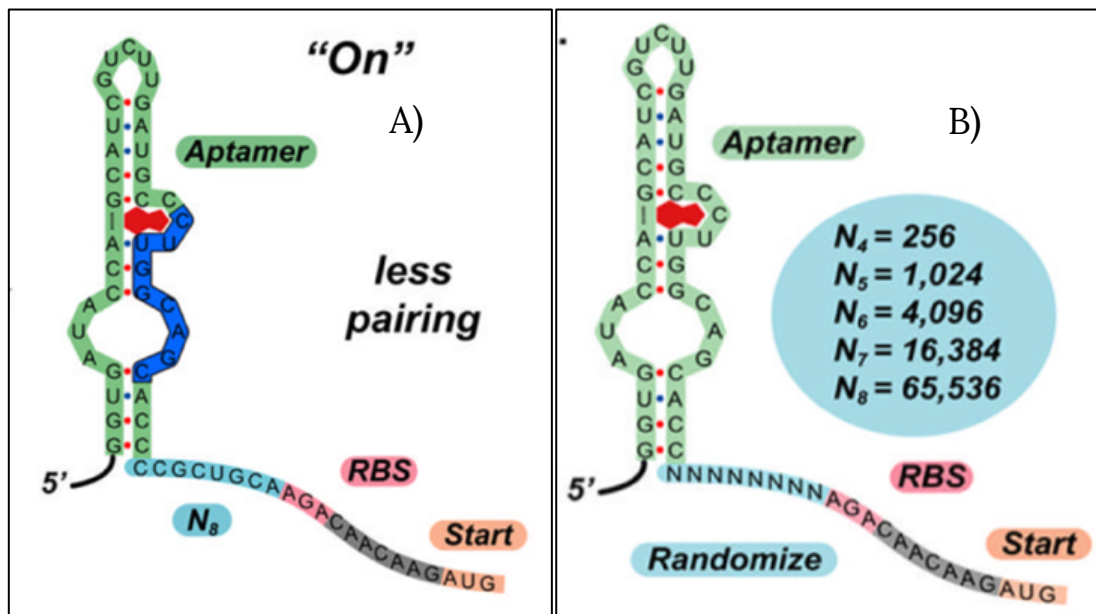


Figure 22. Riboswitch versions.

A) 8.1 version of the riboswitch, which showed the biggest activation fold in Lynch, et al., 2007. B) Structure of the theoretical library obtained after randomizing the 8 nucleotides following the stem of the riboswitch.

Positive selection variants:

Based upon the positive selection plasmid EG\_002+, different combinations of promoter and RBS were tested in order to select for the ideal construct to use in the current experimental design. This would incorporate a combination of



promoter and RBS that gave very low production of *cmk* for a riboswitch in an OFF state (without theophylline), and a higher one for an ON riboswitch (with theophylline). 8 variants were assembled, 6 of them used test combinations, and 2 of them used as controls. The plasmid maps, which can be found in the Supplementary Material and Methods section, were as follows:

- EG\_cI03: (6093 bp, Supplementary figure 2).

The positive control, which contained a T7 promoter and a strong synthetic RBS in control of the *cmk* gene, ensuring its expression, and so, the proper replication of phages, regardless of the presence of theophylline.

- EG\_cI10: (6054 bp, Supplementary figure 3)

The negative control, which contained neither promoter nor RBS in front of *cmk*, so phages wouldn't be able to replicate under any circumstances.

- EG\_cI04-09: (6155 bp, Supplementary figure 4)

The test variants, which contained the previously described pRM promoter followed by 1 of 6 different RBSs of different strengths. These RBSs were known as B0030, B0031, B0032, B0033 (Weiss, 2001), B0064 (Modified from Weiss, 2001), and B00Syn, obtained via the use of the web server version of the RBS calculator (<https://salislab.net/software/> from Salis, Mirsky and Voigt, 2009)

Negative selection variants:

Based on negative selection plasmid EG\_002-. Just like in the case of the positive selection, 8 variants were tested, 2 of them being controls and 6 of them being tests. All the variations are as previously described, except for the plasmids containing the T7 exclusion gene *pifA* instead of *cmk*.

- Positive control/EG\_cI11: 7635 bp, Supplementary figure 5A
- Negative control/EG\_cI18: 7600 bp, Supplementary figure 5B
- Test variants/EG\_cI12-17: 7700 bp, Supplementary figure 5C



Plasmids specific to the Mmfr project

Homologous recombination (HR) plasmids:

A secondary set of homologous recombination plasmids was generated to allow the evolution of a different set of sensors using our phage strategy. Several versions of the plasmid were assembled containing several variants of the gene of interest, the transcriptional regulator *mmfR* (O'Rourke, et al. 2009). These variants included the original gene and two of its variants, named *sgnR* (Sidda, et al., 2014) and SAV2270 (Corre, et al., 2008) and were all based on plasmid EG\_001.

- *mmfR* HR plasmid/EG\_m01: 3726 bp, Supplementary figure 6A.

Containing everything previously mentioned for the HR plasmid, but including a T7 promoter, the lac operator and an RBS in front of the *mmfR* gene (649 bp)

- *sgnR* HR plasmid/EG\_m02 3680 bp, Supplementary figure 6B.

The only difference with the EG\_m01 plasmid is the substitution of the *mmfR* for its *sgnR* variant (600 bp).

- SAV2270 HR plasmid/EG\_m03: 3738 bp, Supplementary figure 6C.

Difference with EG\_m01 plasmid is the substitution of *sgnR* for *avaL1* (657 bp)

Selection plasmids:

As part of the evolution strategy, a selection strategy was also developed. As in previous strategies, these plasmids are based upon EG002, with *cmk* and *pifA* driving the selection. These plasmids contain GFP as a reporter in front the regulatory genes, to indicate when gene expression is happening in the system.

- Negative selection A/EG\_m04: 6953 bp, Supplementary figure 7A.

Containing the original operator *mmyB*, *sfGFP* and *cmk*.

- Negative selection B/EG\_m05: 6953 bp, Supplementary figure 7B.

Same as EG\_m04, except for the mutated version of *mmyB* operator, MARE-4.

- Positive Selection A/EG\_m06: 8502 bp, Supplementary figure 7C.

Containing the original operator *mmyB*, *sfGFP* and *pifA*.

Positive Selection B/EG\_m07: 8501 bp, Supplementary figure 7D.

Same as EG\_m07, except for the mutated version of *mmvB* operator, MARE-4.

### 4.3.2 Experimental procedure for riboswitch evolution:

The initial aim for the project was to introduce the sequence of the SP6 RNA polymerase, which had already been used in directed evolution efforts (Breaker, Banerji, & Joyce, 1994), within the homologous recombination plasmids, downstream from the riboswitch sequence to be evolved. Due to an inability to assemble these constructs correctly, as well as some results from other members of the lab that seemed to indicate this strategy was not as efficient as expected, it was discarded in favour of the  $\lambda$  cl strategy.

After having produced the two riboswitch-containing phage populations, the phages containing the fixed version of the riboswitch were tested against the positive selection strains, *E. coli* BW25113 $\Delta$ *trxA* $\Delta$ *cmk* modified from the Keio collection transformed with the corresponding plasmids. These tests included plaque assays, growth assays, and one-step assays; both in presence and absence of 1.5 mM theophylline in the media; working as an indicator of which strains offered the highest activation fold when the riboswitch is ON, and whether or not the system “leaked”, producing Cmk and allowing the phage to replicate.

Once the positive selection strain was selected, the experiments for the negative selection started, by using Keio *E. coli* BW25113 $\Delta$ *trxA* cells transformed with the plasmid. These cells allowed phages to replicate properly if the riboswitch they contained was OFF in the absence of theophylline. If for some reason it turned ON, this will activate the *pifA* gene, and block replication. The procedures were done in the same fashion as with the positive selection, but due to much clearer results, only some of the assays were made in the case of the negative selection.

After both strains were selected, bioreactor experiments could be properly set up, with phages being present in a “lagoon”, where different strains were added, ideally switching every hour (Times were tentative to improvement). The system flushed the cells before they had time to replicate, but after the phages managed to do so. This meant evolution only affected phages, as cells were not able to maintain any mutations they develop in the system. The lagoon also contained the chemical compound N-methyl-N'-nitro-N-nitrosoguanidine (NG) (Sugimura et al., 1966) to increase the phage's mutation rate, making up for the possibly smaller size of the factual library compared to the theoretical one. The process would go on for several rounds, with samples taken at different time points during the experiment; and the multiple phages will be sequenced in order to observe the differences in the riboswitch, and its impact on infection rates as evolution goes on. A diagram of the bioreactor, as well as schematic version of the procedure are shown in Figure 23 and Figure 24.

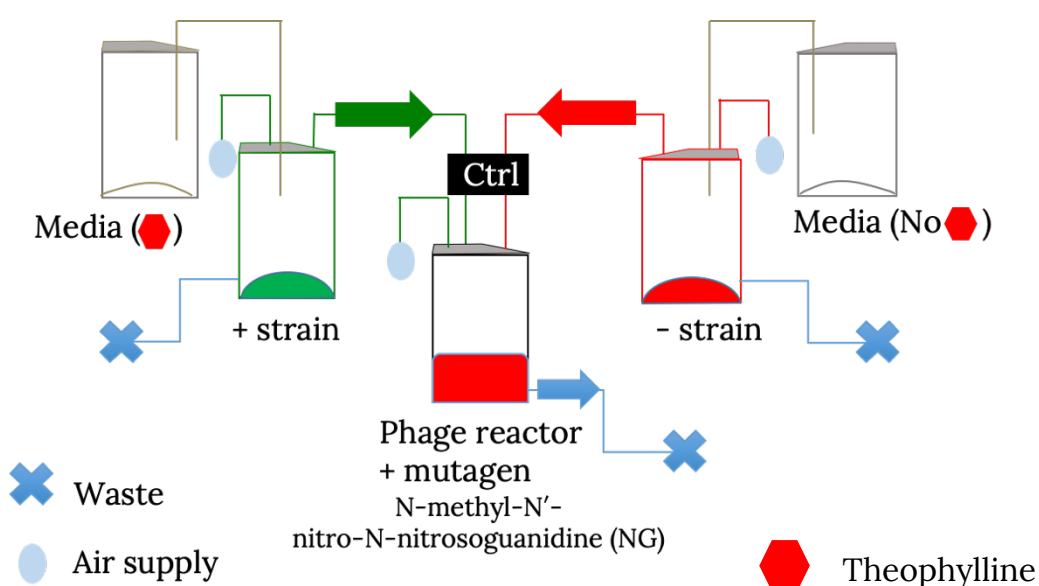


Figure 23. Schematic representation of the original bioreactor.

The figure represents the different elements that form the bioreactor: media that is added to each of the selection strains to ensure constant growth, two cell vats where each of them is grown and their OD monitored, with proper aeration and waste disposal, using chloroform to ensure no cell remains in the pool. Each strain will then be alternatively added to the phage reactor for infection; with a washout time higher than the phages' replication time, but lower than the cells', ensuring mutations are only conserved within the phage population.

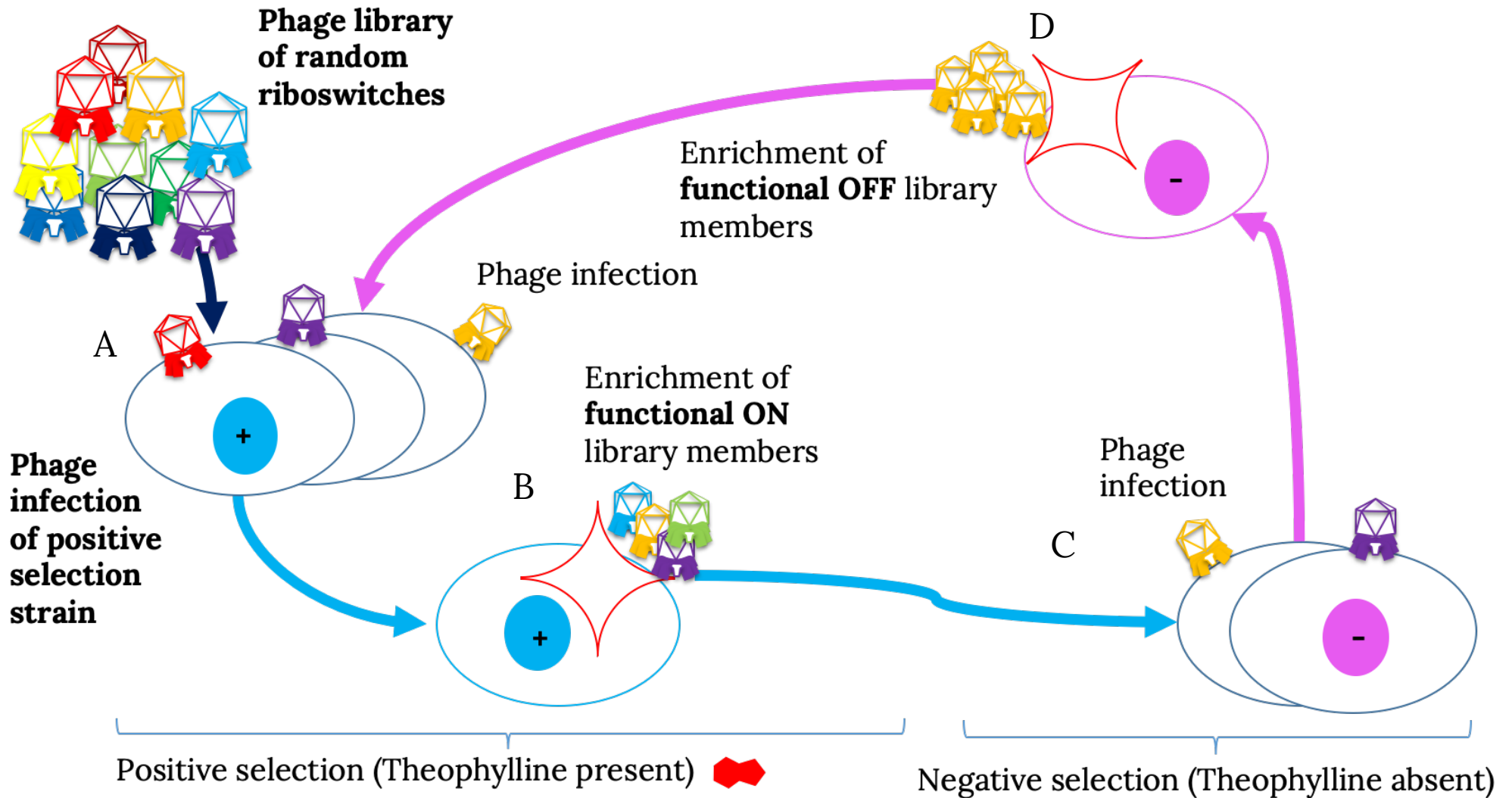


Figure 24. General structure of the full phage evolutionary procedure.

General structure of the originally designed procedure. A) A phage library of switches is used to infect the positive strain, containing a positive selection plasmid (blue +). B) In the presence of theophylline, only phages with working variants of the riboswitch will be able to replicate properly. The system washes out cells before they have time to replicate, ensuring phages unable to replicate within that timeframe are eliminated, as well as keeping phages as the main driver of variability and evolution in the system. C) These surviving phages will then go on to the negative selection step after a time that can be regulated according to the need. D) In this case, the absence of theophylline means that riboswitches should stay inactive, but in order to eliminate constitutively active phages, the *pifA* gene gets transcribed and those phages are unable to replicate properly. The resulting phages will then go back to infecting the positive selection strain after the desired time, and then the system will start once again in 1), progressively forcing the evolutionary pressure towards functional and improved riboswitches over time.

However, due to technical difficulties with the bioreactor arising in the last year of the project, the experiments that had been planned had to be adapted to hand-made individual assays, using a serial dilution approach. In this case, the experiment started with 2 ml of the selection strain being infected at OD 0.2–0.3 with 20  $\mu$ l of the phage library of riboswitches, at an approximate MOI of  $10^{-5}$ . The infected culture was grown for 1 h at 37° C, after which 1 ml was taken out and spun down at 3000 g for 10 minutes; with the resulting supernatant being transferred to a new 1.5 ml tube, 10  $\mu$ l of chloroform added and then vortexed for 15 seconds. This sample was diluted up to  $10^{-12}$ , in order to develop a spot assay and observe the approximate number of phages obtained. These would then be used to infect the negative selection, and the process reiterated for as many generations as wanted. Sixteen rounds of double selection were made for this thesis, with the concentration of theophylline reduced to 1 mM after the tenth one. The process is the same for the *mmfR* project, with the only difference being the regulatory elements in this case.

### 4.3.3 Evolutionary procedure:

Starting from the initial library of random riboswitches obtained after homologous recombination, phages were used to alternatively infect the selection strains, starting with the positive selection. 2 ml of infected cultures at

an OD<sub>600</sub> nm between 0.2-0.3 and an approximate MOI of  $10^{-6}/10^{-7}$ , each time, and put in a shaking incubator at 37°C. One hour after infection, 1 ml of the culture was removed, and the cells pelleted by centrifugation at 14,000 *g*. The supernatant was then transferred to a new tube, where 10 µl of chloroform was added to lyse any remaining cells. These phage samples were then serially diluted 12 times, and used in a spot assay, with 3 samples of 3 µl for every dilution. Once dried, the plates were grown at 37°C overnight, and the plaques counted to calculate the approximate number of plaque-forming units (pfu) in each sample. In some cases, plaque assays were made to ensure a more specific count of the pfu/ml of each sample.

Phage samples were then labelled according to their generation number and stored at 4°C.

### 4.3.4 Next Generation Sequencing preparation

This strategy was designed following the 16S amplicon sequencing protocol from Illumina, using its Nextera XT DNA Library Kit (Illumina #FC-131-001) to prepare the samples so they could be run on an Illumina MiSeq.

PCR was carried out to amplify a 175 bp fragment from the phage genome containing the closest elements present upstream and downstream from the region containing the 8 randomized nucleotides of the riboswitch (Figure 25A). The primers used in this PCR also incorporated sequences necessary to carry out the indexing PCR for the sequencing (Figure 25B). Combinations of primers used in the Index PCR are shown in the supplementary materials, in Supplementary Table 3.

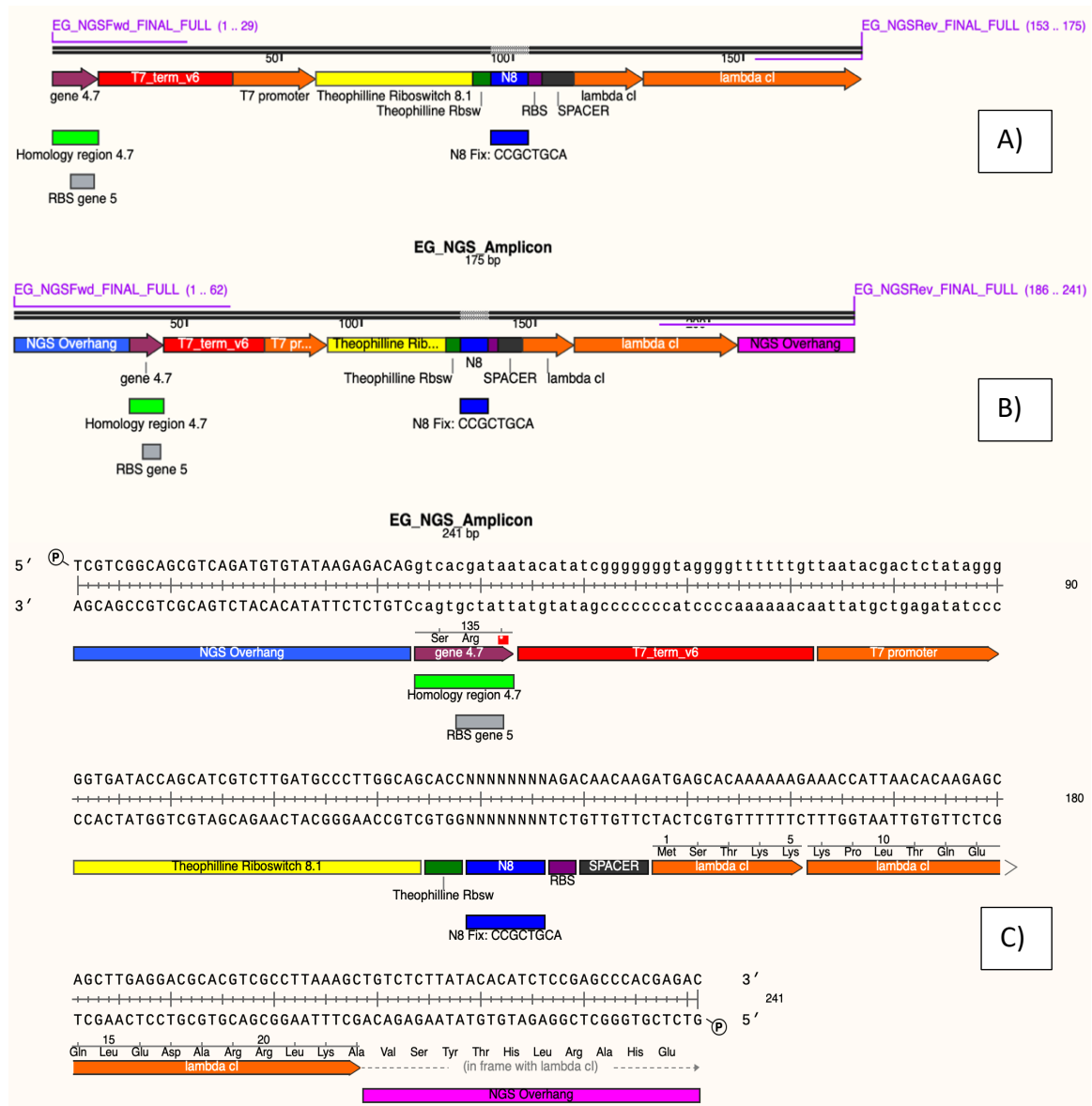


Figure 25. Amplification steps for NGS preparation.

Selected amplicon to be used for the NGS procedures. A) Original amplicon as it is present in the recombined phages' genome, 175 base pairs in size, containing fragments upstream and downstream from the random riboswitch sequence. B) Amplicon obtained after the PCR, with the desired primers having added the required Illumina overhangs for the Indexing PCR. C) Detailed sequence of the final amplicon, indicating the original sequences and the overhangs.

#### 4.3.5 Quantitative PCR assays of samples.

The Illumina NEBNext Library Quant kit (New England Biolabs #E7630S) was used to carry out qPCR to determine the concentration level of each index PCR. Sample preparation was carried out according to the manufacturer's protocol.



The PCR was run for a total of 35 cycles, with the distribution of samples shown on Supplementary table 3.

### 4.3.6 Data analysis of the Next Generation Sequencing results:

After the results from the MiSeq run, Fastq files for each of the 24 different samples analysed in each run were produced using the Illumina FASTQ Generation App, V 1.0.0. Each file contained hundreds of thousands of sequences. These sequences required several different treatments to enable their proper analysis.

The sequence analysis was approached in two different ways:

- Via the use of the AptaSUITE program (Hoinka, Backofen, & Przytycka, 2018) in order to categorise the variation over generations of phages.
- Via the use of several independent Bioinformatic tools to trim, align, and organise the multiple sequences over generations.

AptaSUITE is an opensource software developed specifically as a way of condensing all the bioinformatic workflow necessary to analyse high-throughput SELEX data, all within the same framework. AptaSUITE offers the possibility of recreating an accurate version of the selection process and cluster the different sequences present into each library pool, following their evolution throughout the subsequent generations.

Despite all its positives, several issues essentially stemming from a lack of proper troubleshooting of the software and computational power, this approach was discarded in favour of the more traditional, (albeit slower) use of independent bioinformatic tools to extract the relevant data from the NGS sequences. This analysis, and the corresponding results shown in Figure 33, were made by Carlos Bondia, as stated in the Declaration on page 6 of this thesis.

Due to the lack of computer power, the initial set of analyses were done in a fraction of the total number of sequences, approximately  $10^5$  per sample.

Sequence quality control was performed using the FastQC software (Andrews, 2010). After trimming, all reads corresponding to each sample were aligned to the



initial “parent” sequence (as designed and cloned into the initial vector prior to selection) using a recursive Needleman-Wunsch pairwise alignment algorithm (Rice, Longden, & Bleasby, 2000).

Once aligned, all reads were trimmed from both ends until only the 8 nt-long riboswitch plus 5 bases either side remained. Trimming was performed using base R, by finding the 8 nucleotide long “NNNNNNNN” motif in the parent sequence and extracting this position ( $\pm 5$ bp) for all reads in each sequence. Once extracted, nucleotide proportions were calculated by stacking each position and counting the number of each nucleotide that was informative (not ambiguous), which is considered a Bit. The program doesn’t only count which percentage of each nucleotide is present in every position, as there are many reads showing an ambiguous read or just a gap in these positions. Hence, in order to account for this, it creates a list and only counts the number of times every base is clear, represented as an informative Bit, according to information theory. The entire process of nucleotide counting and plotting was performed using the R package *ggseqlogo* (Bolyen et al., 2019)

### 4.3.7 Evolution and efficiency assessment via plate assays

To quantify the possible evolution of the phage libraries over time, an experiment was devised which would compare the replication efficiency of different phage libraries of riboswitches at different points in the evolution timeline. Both the positive and negative strains were grown at 37°C in 2 ml of LB culture, in the absence and presence of 1 and 1.5 mM theophylline, to see how the phage libraries behaved in each selection step, both when the system was expected to be functioning and also not functioning. All cultures were infected at a MOI of  $10^{-8}$  and grown for 1 h. Such a small MOI was used to allow the phages to replicate for such a period of time without fully clearing the culture. After the incubation 1 ml from each culture was removed and the cells pelleted by centrifugation. The supernatant was transferred to a new tube and 15  $\mu$ l of chloroform added, before the sample was vortexed for 15 seconds to lyse the remaining cells debris.

These samples were serially diluted 9 times each, then used for spot assays with  $\Delta$ trxA-gp5 cells, in order to get a rough estimate of their PFUs. After an overnight

incubation at 37°C, and with such an approximation, plaque assays were done with the optimal phage dilution, allowing for a much more accurate quantification of the resulting PFUs. The PFUs corresponding to each of the conditions were represented in order to compare the phages' efficiency in each of them and were then used to calculate the increase ratio of phages from the initial numbers used for infection.

Taking the PFU values for both active and inactive states in the positive and negative selections allowed for a tentative value named "Activation efficiency" to be obtained, indicating the ratio in which the desired state of each selection produced phages compared to the undesired state of the same selection.

### 4.3.8 Evolution assessment via calculation of the Virulence Index

The selected procedure was the one used by Xie, Wahab and Gill, (2018) and Storms & Sauvageau (2019), by which a value known as the virulence factor was ascribed to different phages in order to compare them to one another.

The virulence index was obtained via a two-step process:

- I. Calculating the local virulence ( $v_1$ ) via the following formula:

$$\text{Equation 1: } v_1 = 1 - \frac{A_i}{A_0}$$

where  $A_0$  represents the area under the growth curve of a phage-less cell culture up until reaching stationary phase, and  $A_i$  the area under the growth curve of a culture at a certain MOI until the same timepoint.  $v_1$  is measured in a range of 0 to 1, with 0 being no virulence whatsoever, and 1 being the maximum theoretical virulence, instantaneous death.

- II. Once the local virulences over a range of MOIs were assessed, they were represented on a curve. The area under that curve ( $A_p$ ) was calculated following the formula

$$\text{Equation 2: } A_p = \int_i^0 v_1 d(\log MOI)$$

with  $i$  being the base10 log of the lowest MOI, which in this case was -6.

- III. That value was then divided by the theoretical maximum value ( $A_{\max}$ ) for the same curve, giving the virulence value for a specific phage on a specific strain. In this case, this maximum value corresponded to 6.
- IV. These values were readily comparable with each other, and were taken both as they were, and also to calculate the “Activation efficiency” for the riboswitches inside the phages, indicating the fold increase between states of the switch in those cases.

The virulence index measurements were made for 5 different phage populations; 3 at different stages of selection, 1 for the original riboswitch-carrying phage, and one for WT T7. Each of these phages was used to infect the positive and negative selection strains under three conditions: 0 mM, 1 mM and 1.5 mM theophylline; in order to assess their virulence under each of those conditions.

All calculations were made using Microsoft Excel 2019 and Graphpad Prism 8.

### 4.4 Results:

#### 4.4.1 Riboswitch evolution

In the first step, considered the positive selection step, the objective was to select for riboswitches that are able to change their conformation and activate the expression of *cI* in the presence of theophylline. The *cI* repressor encoded by the *cI* gene would bind to a specific promoter present in the plasmid within the positive selection strain, and activate the production of the CMP kinase, ensuring the correct replication of the infecting phages. This ensures selection is achieved. Unless the phages carried a working version of the riboswitch, they would not be able to activate the molecular pathway that leads to their correct replication.

### 4.4.2 Determination of best candidate strain for positive selection step:

As previously stated, eight different strains were developed in order to set up the positive selection step of the evolutionary process; and Two of these different strains were controls, but the remaining six needed to be assayed in order to determine which combination of pRM promoter and RBS would work best in this case. This was decided via the use of growth, plaque, and one-step assays, both in the presence and absence of 1.5 mM of theophylline; and a comparison of the obtained results for each strain. The combination that showed the highest activation fold between absence (OFF state) and presence (ON state) of theophylline would be selected as the one used in the evolutionary process.

To set up this procedure, the eight different plasmids were designed, constructed and transformed into *E. coli* BW25113  $\Delta$ trxA $\Delta$ cmk cells, each of them carrying the same elements except for the promoters in front of the *cmk* gene, to be used for selection. All were put up against the phage population containing the fixed version of the riboswitch, originally carried by EG\_cI01, in order to determine the efficiency of the regulation. The sequences for the different RBS were taken from the iGem parts website ([parts.igem.org](http://parts.igem.org)), originally mentioned in (Weiss, 2001). These RBS were known as B0030 to 33, but also included B0064, which is based on the original sequences, and a synthetic RBS obtained using the RBS Calculator software from the Salis Lab (Salis et al., 2009)

Plasmid EG\_cI03 was used as a positive control, containing the *cmk* gene under the regulation of a T7 promoter and a strong RBS. This ensured that whenever a phage infected the cell, the gene would be produced and the resulting phages would replicate without further issue. Plasmid EG\_cI10 was used as a negative control, with neither promoter nor RBS, ensuring no replication of the phages could occur regardless of the conditions.

The efficiencies of these strains in phage replication were tested with via killing assays, measuring the OD values under different conditions and comparing between them in order to find the most suitable candidate. Plaque and spot

assays were also carried out, where the number of plaques present in a petri plate is directly related to the PFU of the phage sample. All these assays were done both in the presence and absence of 1.5 mM of theophylline; and the obtained results from each strain were then compared. The combination that showed the highest activation fold between presence (ON state) and absence (OFF state) of theophylline would be selected to be used in the evolutionary process.

Figure 26 shows the results obtained from the killing assays under these conditions. For each of the 8 strains tested, we can see the two growth curves at OD<sub>600</sub> nm, corresponding to the strain growing in the absence or presence of theophylline with a certain value of T7 phages containing the fixed version of the riboswitch. These results indicate that among the tested strains, only two of them show enough difference due to the presence of theophylline to be considered relevant, something that can be blamed on the leakiness of the pRM promoter, which allows for some gene expression to occur in all cases.

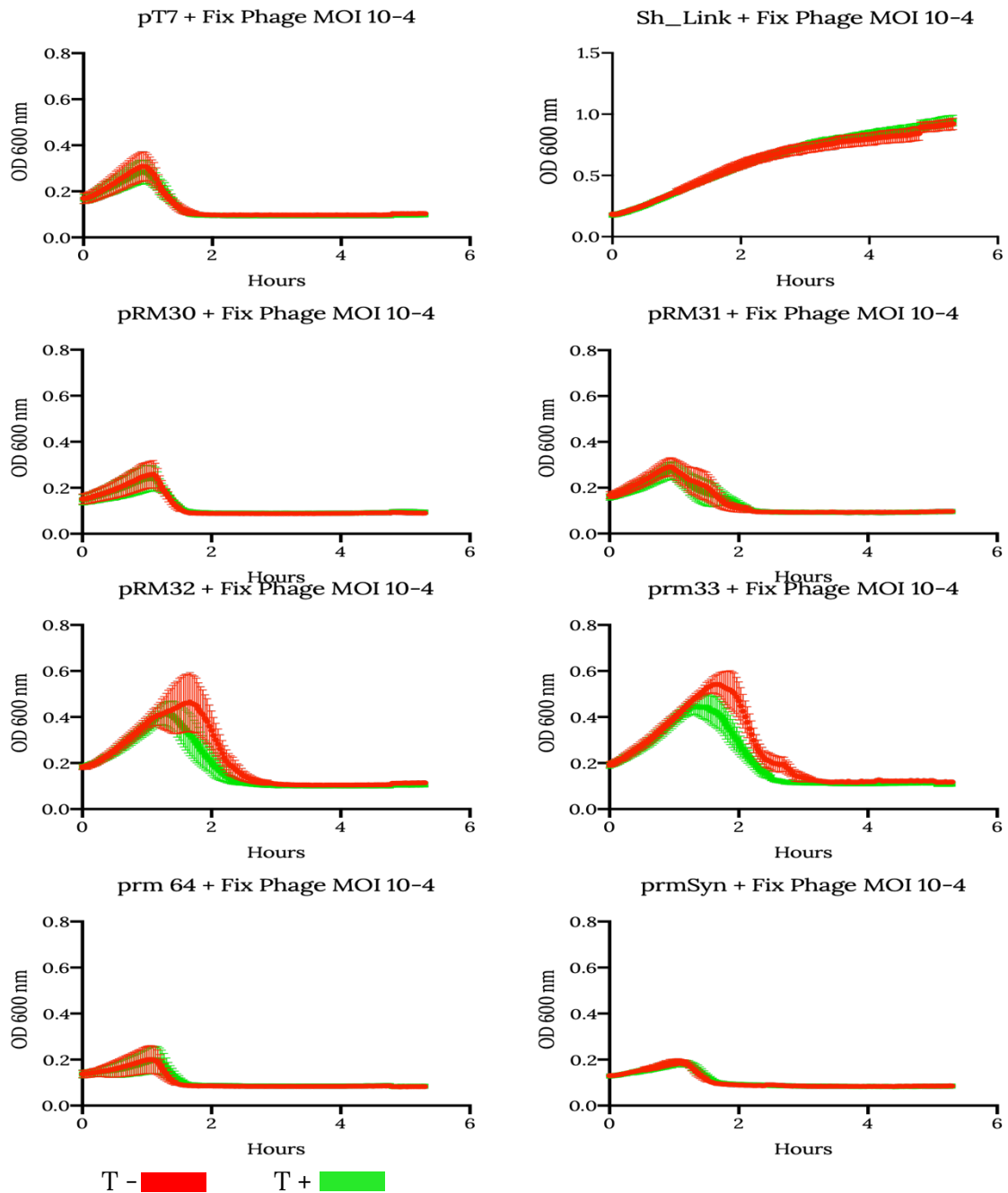


Figure 26. Results from cell killing assays in positive selection strains.

Graphic representation of OD<sub>600</sub> over time, showing the killing of each positive selection strain, including controls, by Fixed Riboswitch-containing phages at a rate of 1 phage per every 10<sup>4</sup> cells. T- represents the case where theophylline is absent, while T+ represents the case where theophylline is present at a concentration of 1.5 mM. The positive control, (A) shows no difference in the killing rate regardless of the presence of theophylline; while the negative control (B) shows the opposite reaction, with no killing in any case. Each of the different test strains (C-H) shows different killing curves, which can be associated with the strength of the particular RBS present in front of the *cmk* gene, as well as its

leakiness, based on the difference between the two curves present in each graph. The bigger the difference between the presence or absence of theophylline, the more efficient the riboswitch's activation is in that particular case. Error bars in all graphs represent the Standard Error of the Mean (SEM)

Once the number of strains to be tested was lowered, the assays known as One-step assays took place, as explained in Figure 10 in the Materials and Methods section. These assays are essentially plaque assays, where the number of plaques is counted and correlated to the number of phages used in the experiment to determine an approximate concentration of phages per ml. One-step assays are more specific in that measurements are taken every 4 minutes over 40 minutes, in order to see the change of phage numbers over time. The result is then compared within the different conditions in a more quantitative manner than that of OD measurements. For this assay, only four strains were tested, including the positive and negative control (Figure 27).

These results show that even though there is a difference between the presence and the absence of theophylline in the tested strains, such a difference is higher in the case of the strain containing the combination of pRM promoter and RBS B0033, making it the best candidate to use for the positive selection step of the evolution procedure.

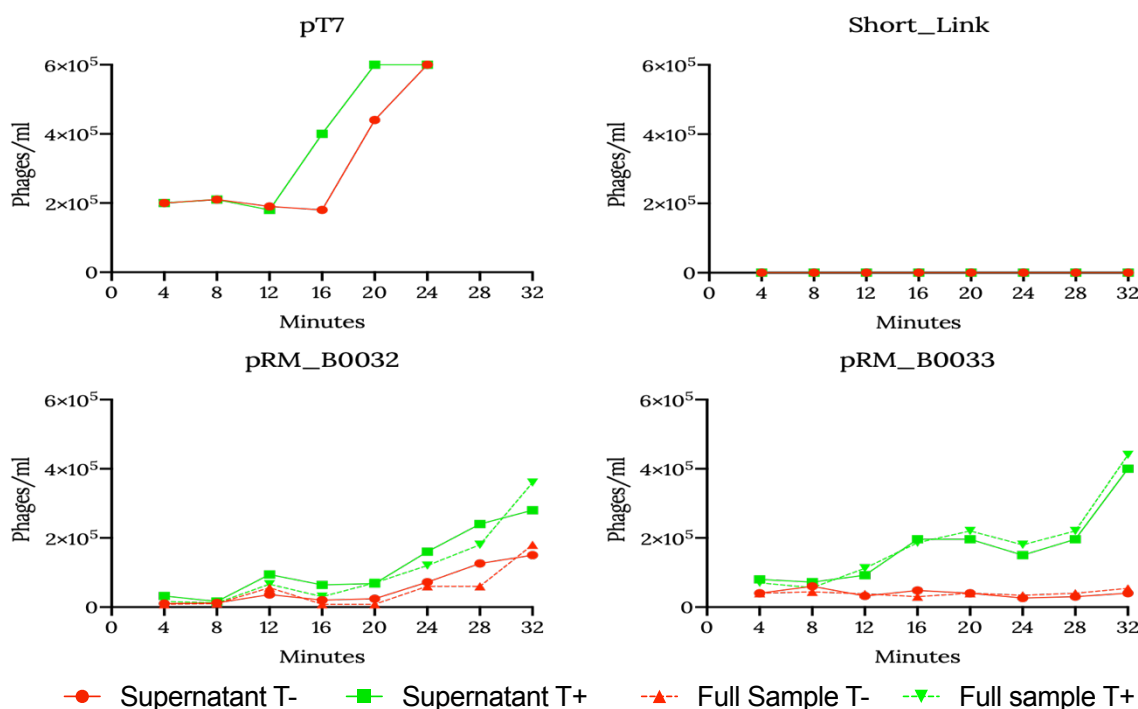


Figure 27. Results from One-step assays in positive selection strains.

Graphic representation of phage production for different cell treatments in the presence (T+) or absence (T-) of theophylline. As explained in Materials and Methods, in one set of samples cells are spun down and the supernatant taken (Supernatant label), while in the other, chloroform is added to the whole sample (Full sample label) both with and without the presence of theophylline. A) Phage production for the positive control strain, with *cmk* being regulated by pT7. B) Phage production from the negative control strain, with no promoter nor RBS in front of the *cmk* gene. C) Phage production for a test strain containing the combination of pRM promoter and RBS B0032. D) Phage production for the test strain containing the combination of pRM promoter and RBS B0033.

#### 4.4.3 Determination of best candidate for negative selection step:

Once the first step was established using the control phage, the same process was repeated for the negative selection step. If the positive selection was designed to select those versions of the riboswitch that turned ON in the presence of theophylline, this one selected for those that remain OFF in its absence. Due to the switching nature of the riboswitch, such a step is essential in order to eliminate elements within the population that “cheat” their way through by being constantly on. In such cases, where the switches are still on even in the



absence of the activator, *ci* will be produced, bind to the pRM promoter and activate the expression of the gene *pifA*, which codes for the T7 exclusion protein PifA. The actual molecular mechanism by which the exclusion is happening is not fully clear (Cheng et al., 2004), but when such a protein is expressed in a cell at the same time as T7 genes 1.2 or 10, the cells shuts down, blocking the replication of phages within it (Schmitt & Molineux, 1991).

In order to have an efficient negative selection, a similar set of assays to those developed for the positive selection had to be developed, in order to find the best possible candidate for phage replication when using the “Fixed riboswitch” phage population. Once again, cells were transformed with eight different variants of the selection plasmid, EG\_cI11 to EG\_cI18. EG\_cI11 contained *pifA* under the control of the pT7 promoter, ensuring a constitutive production of the protein, while EG\_cI18 consisted of *pifA* without any of upstream regulatory element, diminishing the production as much as possible. The test strains, EG\_cI12 to EG\_cI17, all contain the same combinations of pRM promoter and RBS that were tested in the positive selection, but in this case both the regulated gene (*pifA*) and the genetic makeup of the cell were different.

Cells in this case were *E. coli* BW25113  $\Delta$ *trxA* from the Keio collection, unlike the *E. coli* BW25113  $\Delta$ *trxA* $\Delta$ *cmk* cells used for the positive selection, given that here *cmk* expression does not regulate phage selection, and thus is constitutively expressed so phages can replicate appropriately.

Just like in the case of the positive selection, an initial set of killing assays was carried out to assess the replicating efficiency of the fixed riboswitch (Lynch et al., 2007) phages in the different strains. In this case, however, the results were much more straightforward, as seen in Figure 28, since only one of the test strains (containing *pifA* under the control of RBS B0033) showed any indication of phage killing. All other strains seemed unaffected by the different phage concentrations present in the media, rendering them unusable in the selection procedure.

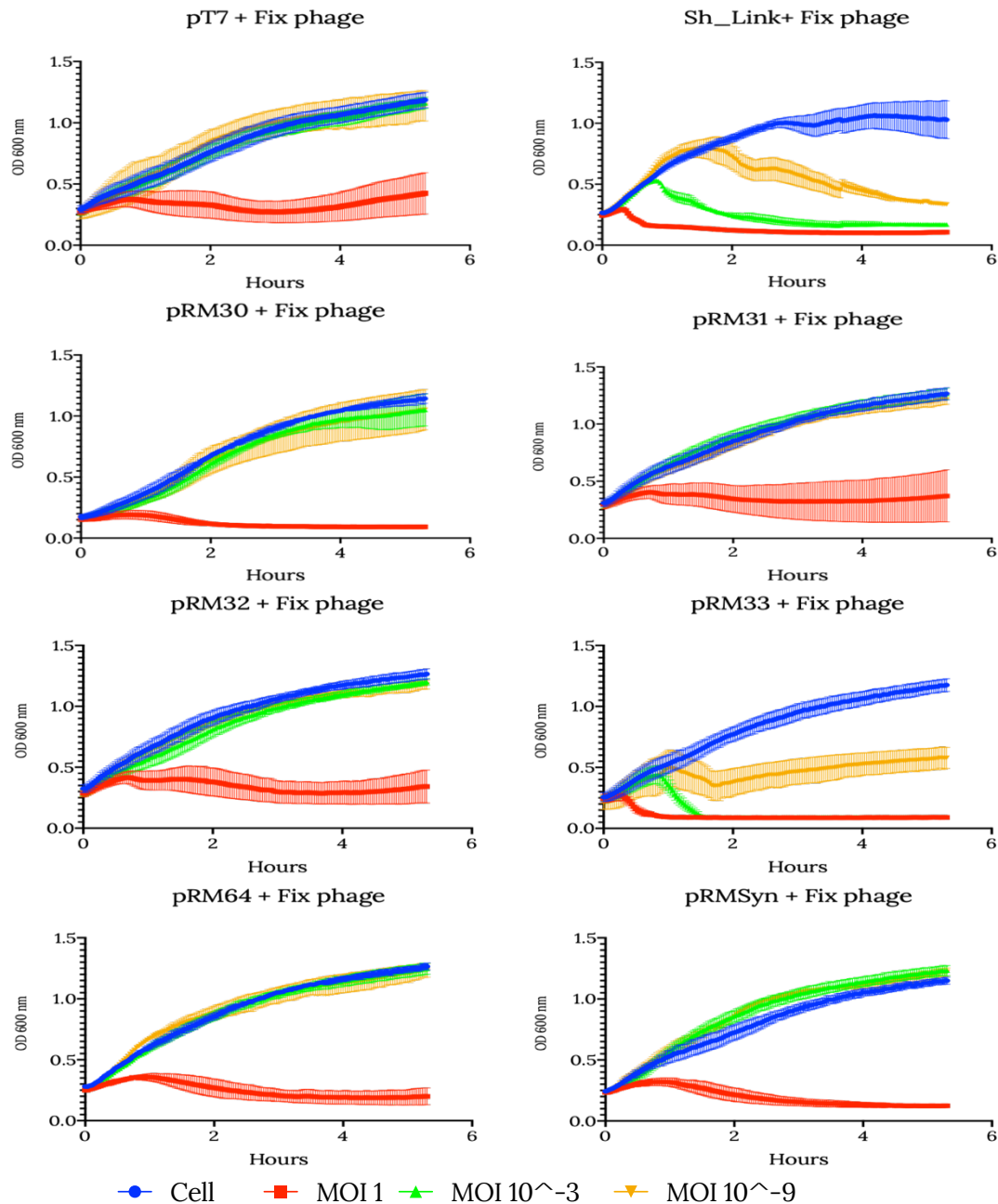


Figure 28. Result from cell killing assays in the negative selection strains.

Graphic representation of OD 600 over time (in hours), showing the killing of each negative selection strain, including controls, by the Fixed Riboswitch-containing phage at different multiplicity of infections, including MOI 0/No phage (Cell), MOI 1, MOI  $10^{-4}$ , and MOI  $10^{-9}$ , all in the absence of theophylline. Positive control (A, with a pT7 controlling *pifA* expression) shows no killing, while negative control, (B) with no production of *pifA*, shows cell killing. Each of the test strains has a specific set of killing curves, but all of them show phage killing only at the highest MOIs, except for the combination including RBS B0033, which indicate the influence of phages is felt even at very small MOIs. Error bars represent the Standard Error of the Mean (SEM).

Based on these results, it was deemed unnecessary to do one-step assays in this case to compare the number of phages being produced in each strain. However, the same data was used to calculate the effective growth rate for each strain in the presence of phages. This measurement refers to the difference in OD in a cell culture between two different timepoints and is calculated as the slope of the straight line between point 1 and point 2 in a semi-logarithmic plot, resulting in this equation:

$$\text{Equation 3 Effective growth rate} = \frac{\ln \frac{OD_2}{OD_1}}{T_2 - T_1}$$

with the result being expressed in  $h^{-1}$ . Based on the 2011 Miller lab protocol (<http://miller-lab.net/MillerLab/protocols/general-bacteriology/calculating-growth-rate/>), based on (Neidhart, Schaechter, & Ingraham, 1990).

This equation was used to calculate the growth rate at the steepest point of the slope for the strain containing the pRM promoter and RBS B0033 combination under multiple MOIs; and then for the same interval of time in the remaining strains, in order to compare them to one another in an effective manner, as shown in Figure 29, where all test strains show positive values for growth rate except for B0033, confirming it is the only case where phages are able to actually replicate and lyse the cells.

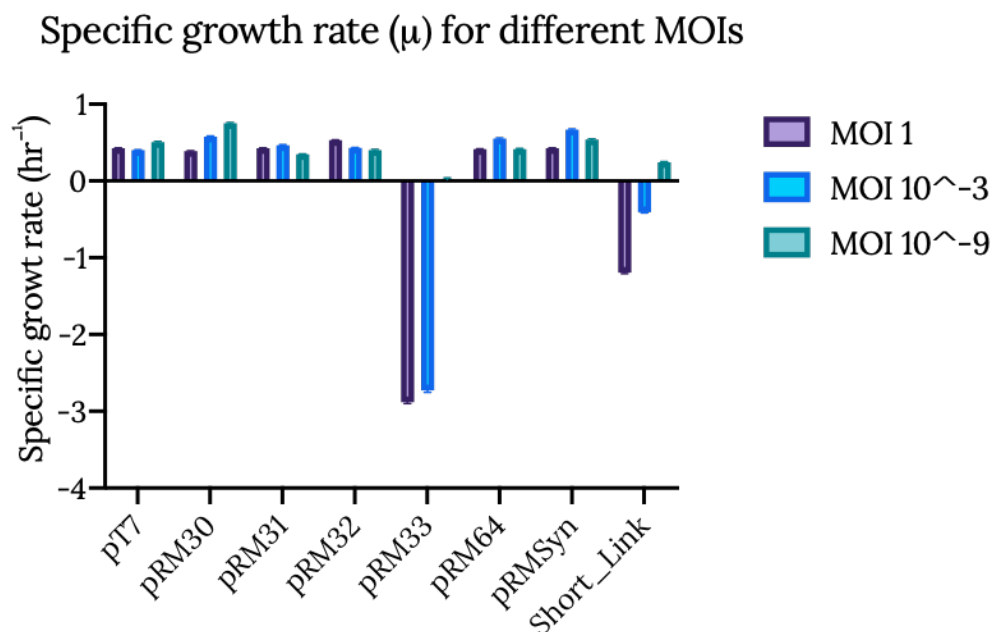


Figure 29. Specific growth rates of negative selection strains.

Representation of the effective growth rates at different MOI values for negative selection strains during the same time period (3000 to 4200 sec) after the start of the experiment. The time period was determined by calculating the steepest slope for the problem strain and then calculating the slopes corresponding to that same period in all other strains. Save for the negative control, where no pifA is produced at all, only the combination of pRM and RBS B0033 show some cell killing, confirming the suitability of that strain for negative selection. Error bars represent the Standard Error of the Mean (SEM).

This confirmed that the specific combination of pRM and RBS 33 was the best candidate to be used for the negative selection process. One further experiment was required to assess the efficiency of the system by comparing its behaviour in the presence and absence of theophylline. The procedure was the same as that carried out for the positive selection, taking samples every 4 minutes for a period of 40 minutes and then seeing the increase of phage particles as infection proceeds. These assays compared the results between the presence and absence of 1.5 mM theophylline; however, the compound would never be present in the system in this case. This assay was made in order to confirm the differences in the system's activation levels between both conditions. The results confirmed that there was indeed a difference between the two conditions, but much smaller than the one showed for the positive selection (Figure 30).

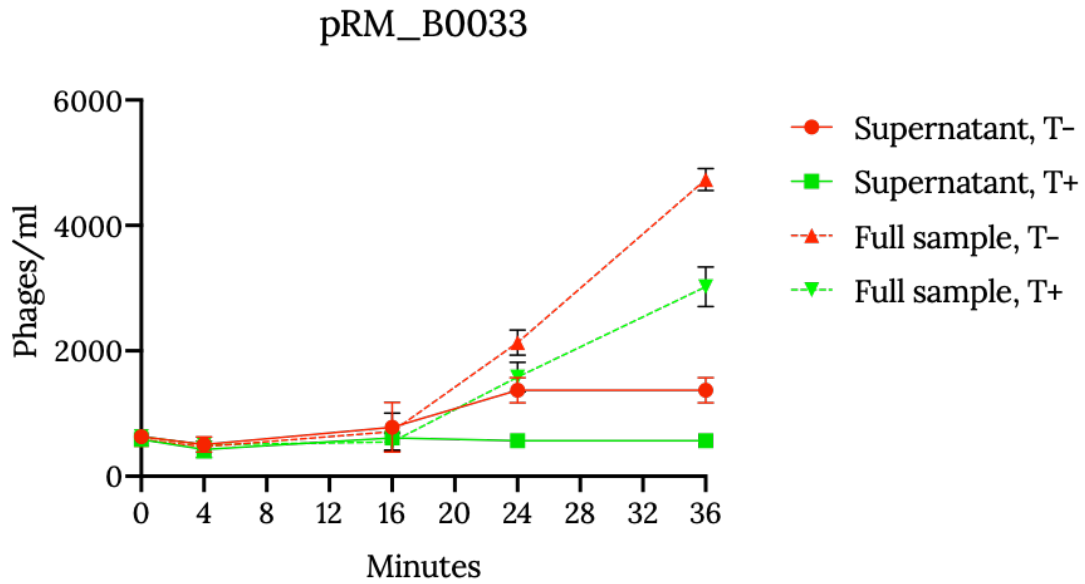


Figure 30. One-step assay for the negative selection step.

Representation of the result from the one-step assays for the negative selection strain containing the pRM promoter followed by RBS B0033, under different conditions and treatments, as described in Materials and Methods, as well as both with and without the presence of theophylline. The graph shows a difference in the cases where only the supernatant were used to infect, indicating a slight difference in levels of phages present in the supernatant, but more so in the case of full samples treated with chloroform, where a more obvious difference can be seen between the two conditions. All of this indicates a bigger phage production in the absence of theophylline. As in previous cases, the error bars here were obtained by calculating the SEM.

#### 4.4.4 Full evolutionary procedure

Having identified strains for both the positive and negative selection processes, the system could now be tested as a whole and its efficiency as a directed evolution strategy established. The original plan for the setup of this experiment was to use the laboratory's self-developed bioreactor, in a process similar to the one presented in Carlson, Badran, Guggiana-nilo, & Liu, (2014). However, the bioreactor suffered a severe malfunction which rendered it unusable therefore, classic plaque and spot assays were used as an alternative approach.

The procedure was done using the previously described spot assays to roughly assess the PFU of each phage generation, so as to keep a similar number of infecting phages between every selection iteration. There were cases over the procedure in which the spot assays were inconclusive, and so, in such a case, both the pre-existing sample and a newly made sample were tested in order to keep going.

So far, 16 full cycles of evolution have taken place, which translates to 32 different selection steps, some of them with more than one replicate. It is important to indicate that all positive selection steps until the 10<sup>th</sup> generation had a concentration of 1.5 mM theophylline, but from that point onwards, the concentration was reduced to 1 mM theophylline, in order to further constrict the riboswitch selection. It was then considered appropriate to test whether some changes to the population of random riboswitches had happened over the course of the procedure.

Preparations for a Next Generation Sequencing assay were undertaken, with the objective of discerning whether the number of riboswitch variants had changed over generations, and give indications of evolution within the populations. Sequencing was carried out using a standard Illumina protocol for amplicon sequencing and an Illumina MiSeq, the objective of the assay was to determine the following:

- 1) Was the original riboswitch plasmid library obtained after bacterial transformation biased in some way? If so, how many of the expected sequences did it contain, and which were missing?
- 2) How many different riboswitch versions were lost, if at all, in the process of homologous recombination to obtain the phage library of riboswitches?
- 3) Has there been any decrease in the number of riboswitch variants present in the libraries as selection proceeded?
- 4) Is there any indication of directed evolution of the riboswitch sequence as generations passed?

A total of 24 different phage libraries were prepared for each sequencing run. The samples included for library preparation incorporated the original plasmid library of riboswitches, the original homologous recombination phage population and the furthest generations present at the point of analysis. All necessary preparations for the experiments are shown in their corresponding subsection. The contents and sequence of the amplicon obtained after the first PCR are shown on Figure 25, and the indexing primer combinations are shown on Supplementary figure 3 of the Supplementary Materials section.

The first sequencing run resulted in no data, due to problems with the sequencing procedure. Therefore, a qPCR assay was carried out to determine the presence of the specific tags in the samples. The Illumina NEBNext kit for library quantification (#e7630) (<https://international.neb.com/products/e7630-nebnext-library-quant-kit-for-illumina#FAQs%20&%20Troubleshooting>) is specifically designed to detect the tags needed for the Miseq run. The sample distribution for this technique is described in the Supplementary Materials section, and the resulting amplification curves confirmed the presence of the fragments in the samples. ()

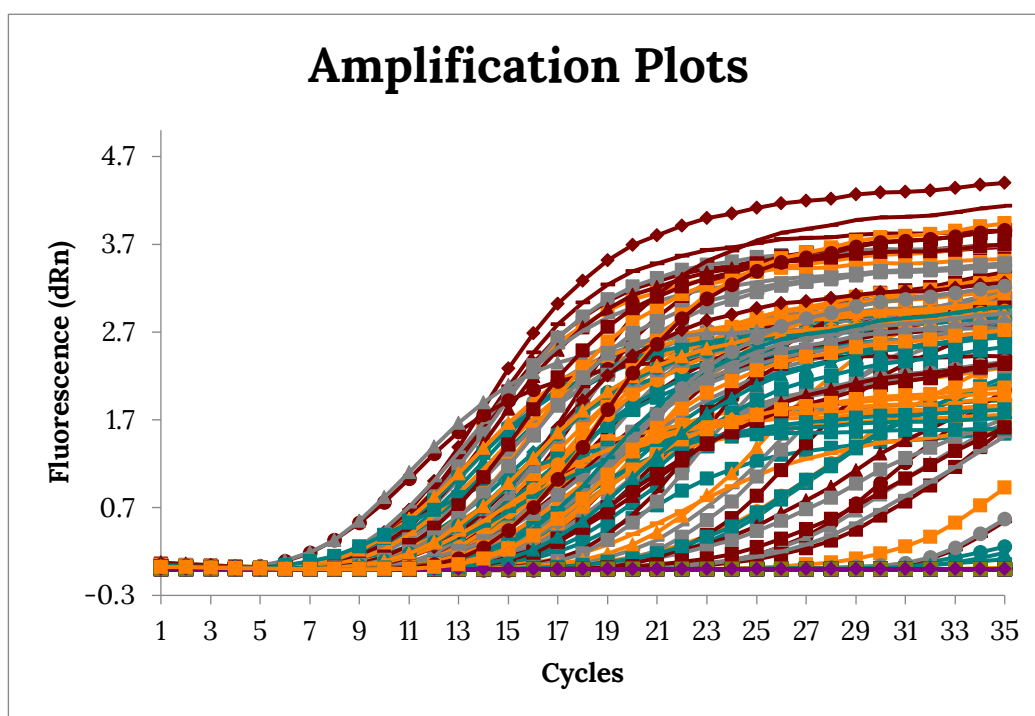


Figure 31. qPCR amplification for Next Generation Sequencing



Results from qPCR assays using NGS-prepared library samples. Thirteen different libraries were run as part of the assay, at two different dilutions, with three repetitions per dilution. Since the qPCR oligos used are specifically designed to only bind to the specific tags used to prepare the NGS samples, the increase in fluorescence confirms the presence of the tags, and thus affirms the validity of the samples for the MiSeq run.

Discussion with the manufacturer pointed towards an unexpected issue with the sequence of the original set of primers used for the first PCR procedure; which prompted a repetition of both this and the indexing PCRs for the samples with a new set of primers. The result from running this final set of indexing PCRs on a 2% agarose gel (due to its small size) shows the band at the corresponding size, between 200 and 300 bp, confirming Indexing amplifications worked. (Figure 32)

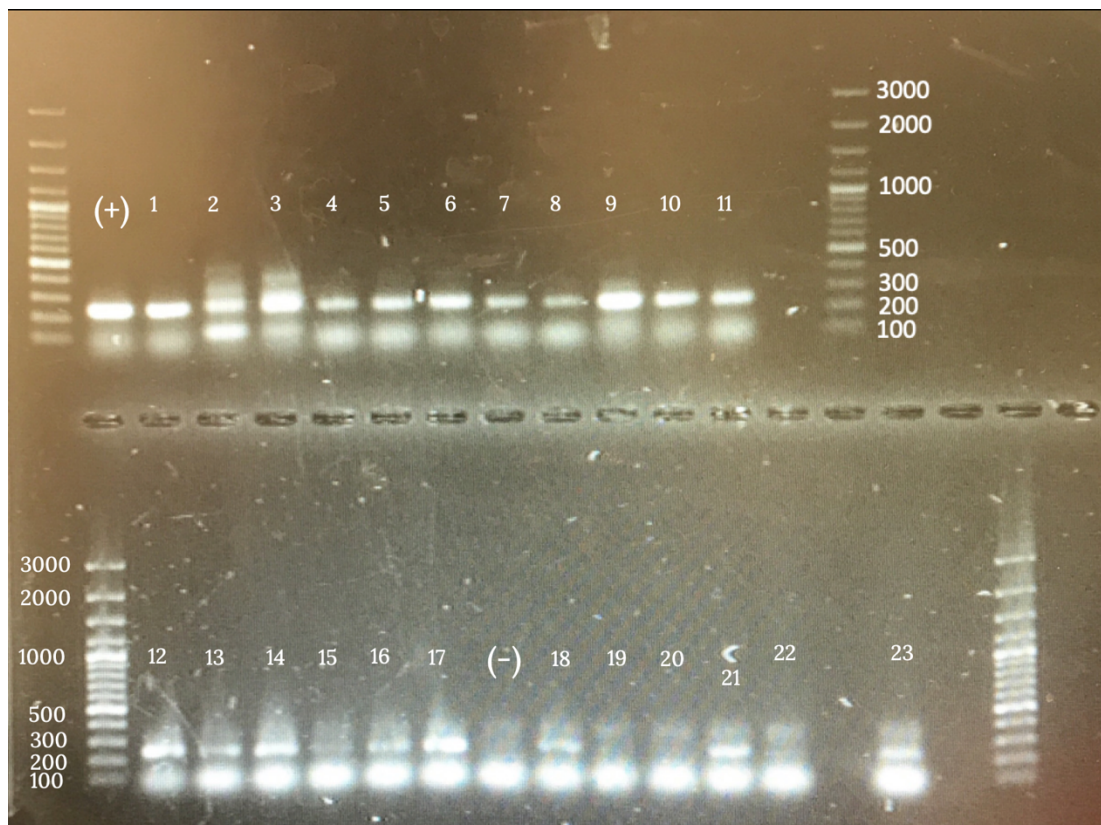


Figure 32. Agarose gel of Indexing PCR.

Agarose gel electrophoresis results of indexing PCR along with a 100 bp plus ladder, showing the size of the fragment at the expected range of over 200 bp and under 300. The thicker, brighter band at the lower part of each lane can be assumed to be a primer dimer, occurring when the primers used in the PCR bind within themselves.



This library resulted in a successful MiSeq run, that yielded a total of 2.58 Gbp, which, given the size of the fragment, roughly correspond to  $10^7$  sequences. These sequences were trimmed, aligned, and grouped, in order to determine whether there is a change in the variation within the riboswitch library over several generations of selection and whether that could qualify as possible evolution.

Due to the sheer amount of data, tackling such an analysis was a very elaborate procedure, and different approaches were taken, to various degrees of success. In order to get a preliminary idea of what each of the libraries held, a small random sample of  $10^5$  reads were taken from each of the 24 libraries and the sequence variation in each set assessed. This number was considered appropriate as it was large enough to encompass the sequence library at its maximum variety (65536 bp), and small enough for the data to be managed appropriately. These results are shown in the form of diagrams representing the 8 random nucleotides flanked by 5 fixed nucleotides on both sides. The larger the size of a particular nucleotide, the more represented that particular nucleotide is in that position.

This was done for multiple libraries throughout the evolutionary process, each step marked as cmk or pifA depending on whether it corresponded to a positive or negative selection step, respectively; and a number to indicate the generation since the beginning of the project. These two diagrams, (one set for each successful MiSeq run) can be seen in Figure 33, showing the changes in sequence distribution over generations, with the corresponding library information in , and indexing primer combinations present in Supplementary Table 2.

## Recombination-based evolution strategy

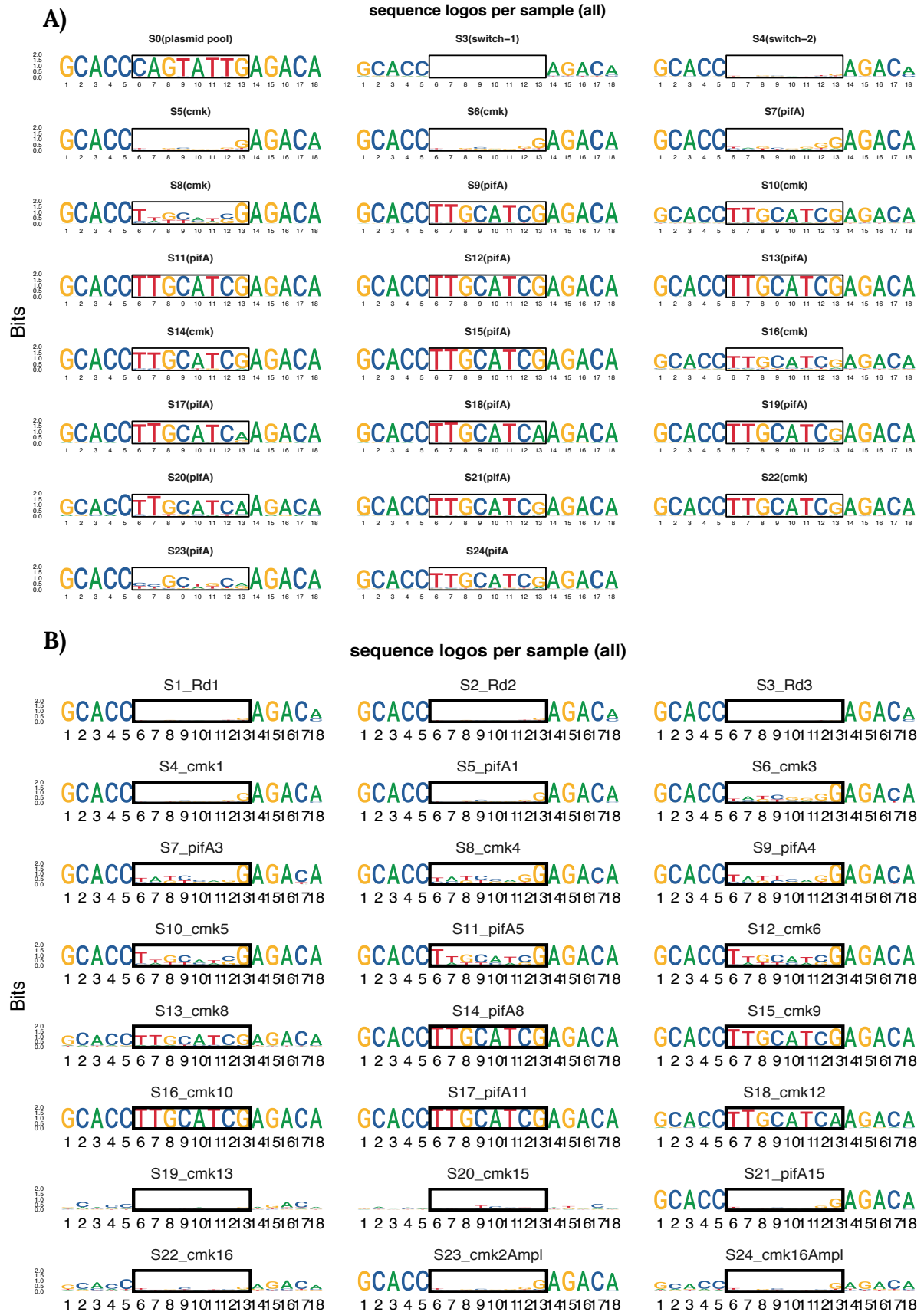


Figure 33. Riboswitch sequence evolution throughout generations.

Diagrams representing the sequence variation in the samples for each of the libraries. A) represents the first successful run, while B represents the second one. Information regarding the samples can be found in . The bases included within the black square correspond to the randomised riboswitch sequences. The larger a nucleotide (AGCT) is represented in the diagram, the more represented it is in the sample population. This explains the constant size of the flanking regions, with some exceptions, probably due to some sequencing issues. In libraries from the beginning of the selection process (Switch-1 and 2 in A, S1, S2 and S3 in B) we can see no prominence of any sequence in particular, exemplified by the empty square, and we see a progressive increase over time towards one dominant sequence, which gains representation early in the process. The sample used for the first library (Plasmid pool/S0) representing the original plasmid library, was wrong due to it probably being the amplification from a single clone, not a full library. Another proof of this is that it shows a predominant sequence at the very beginning of the selection process, which does not correspond to subsequent libraries. The reason why S23 in A, as well as S19, 20, 21 22 and 24 in B shows a much higher variation was unclear here, prompting the analysis of the full data sets, rather than just a fragment of them. A more thorough description of the process can be found in the Materials and Methods section.

Table 7. NGS sample nomenclature.

Correspondence between the generation number of the different analysed libraries and the nomenclature used in the sequence variation diagram. cmk or pifA respectively indicated whether the sample corresponded to a positive or negative step in the selection, and the number indicated the specific generation. Some generations were repeated due to low PFU values, and thus tested to determine any possible correlations between this and the riboswitch sequence. “Samples A” correspond to the samples from the first successful sequencing, and include a wide margin of sequences, from generation 1 to generation 14. Based on these results, the second run included samples indicated under “Samples B”, with samples used to fill in the gaps from the first set, and show consecutive generations until the riboswitch sequences became fixed.

SAMPLES A	NUMBER	SAMPLES B	NUMBER
PLASMID Pool	S0	Rd Switch 1	S1
Fix Switch	S2	Rd Switch 2	S2
Rd Switch 1	S3	Rd Switch 3	S3
Rd Switch 2	S4	cmk1	S4
Cmk1	S5	pifA1	S5
cmk2	S6	cmk3	S6
pifA2	S7	pifA3	S7
cmk5	S8	cmk4	S8
pifA6	S9	pifA4	S9
cmk7	S10	cmk5	S10
pifA7	S11	pifA5	S11
pifA9	S12	cmk6	S12
pifA10	S13	cmk8	S13
cmk11	S14	pifA8	S14
pifA11	S15	cmk9	S15
cmk12	S16	cmk10	S16
pifA12	S17	pifA11	S17
pifA12B	S18	cmk12	S18
pifA13	S19	cmk13	S19
pifA13D	S20	cmk15	S20
pifA13B_8-11	S21	pifA15	S21
cmk14	S22	cmk16	S22
pifA14_6-11	S23	cmk2 Amplif.	S23
pifA14	S24	cmk16 Amplif.	S24

As can be seen in the figures, starting from a point where no sequence is overrepresented (excluding the wrong “Plasmid Pool” sample, corresponding not to a library, but a single clone due to a mistake), the variation within each phage population seems to decrease over the course of the evolutionary steps, as one

single sequence has a progressively prominent presence. This could be taken as confirmation of evolution throughout the generations, but in order to assess so, a different set of assays had to take place.

Several spot and plaque assays took place to compare the efficiency of replication of phage libraries at different points of the evolutionary process. Samples from generation two were set up against samples from generation sixteen, both under the presence and absence of theophylline, and at the same multiplicity of infection, to assess whether there had been any changes in the efficiency with which they replicated within the positive and negative selection strains over the same period of time.

These evolution assessment assays were detailed in the chapter's materials and methods, apart from the general spot and plaque assays described in the general methods. However, this procedure had a significant flaw, namely, the fact that the results were not comparable. Despite multiple repetitions, the variation within samples was simply too big to consider the assays replicates of each other. Possible reasons for this are considered as part of the Discussion chapter.

Because of this, the values of virulence index were calculated, using different phage populations, according to the described method in materials and methods. The selected phages corresponded to the recombinant phage used as a control for the selection process, phage populations from generations 2, 5, and 16; and a WT T7 phage population, lacking a riboswitch in its genome. All populations were tested in all previously described conditions.

A standard set of results for one of the calculations, showing all the different steps, from the OD<sub>600</sub> curves in the presence and absence of phages to the regional virulence values; can be seen in Figure 34. This set of data was obtained after infecting the positive selection strain with phages corresponding to the 16<sup>th</sup> generation, meaning phages that had undergone each selection 16 times in total.

## Recombination-based evolution strategy

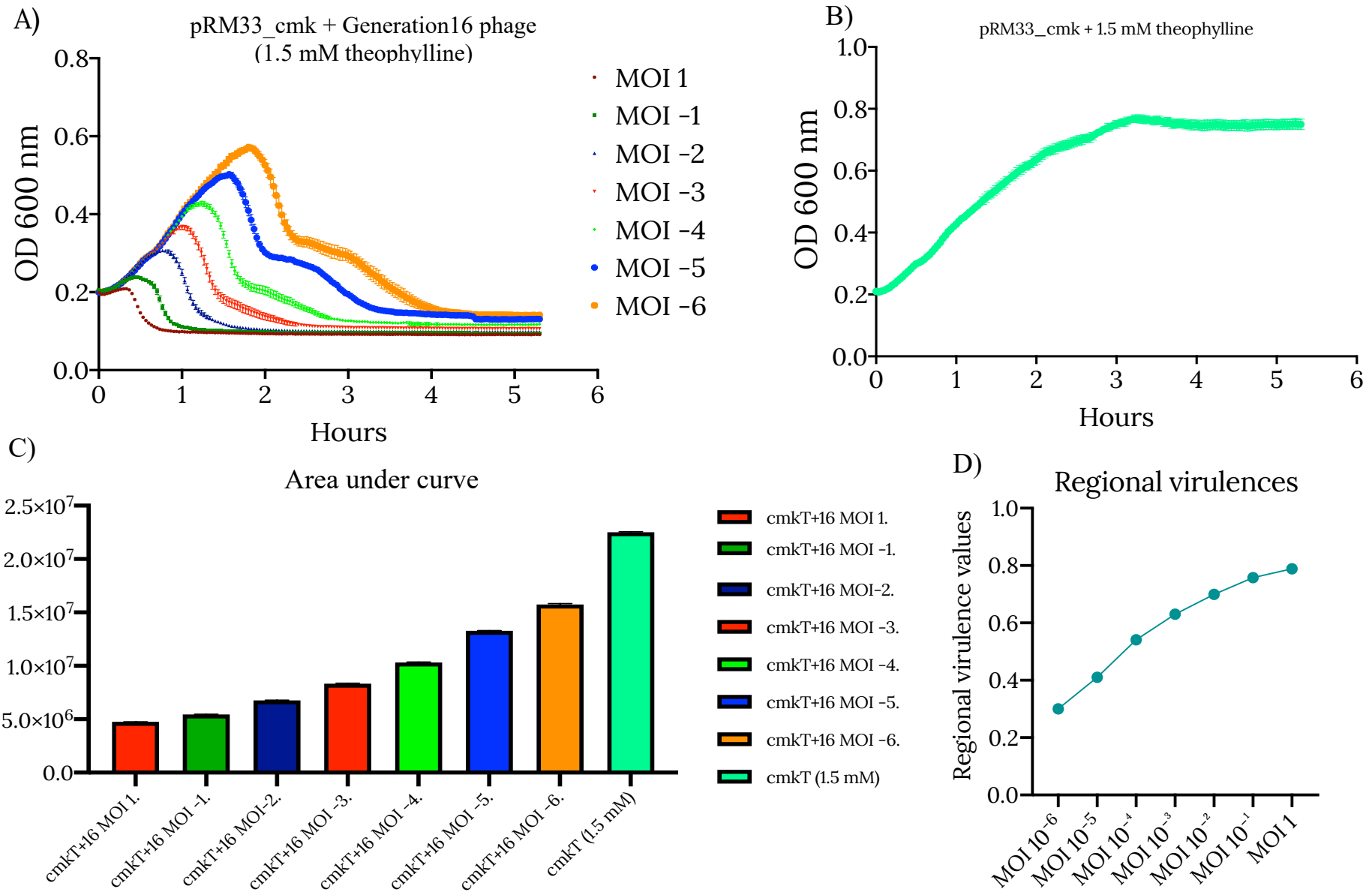


Figure 34. Virulence index example data.

Standard set of data used in order to obtain the virulence index for a phage population infecting a specific strain. A) Killing curves of the cell cultures where phages at different MOIs have been added. B) Growth curve of the cell strain without phages. Notice the point where it enters stationary phase, where the integration of the curve finishes. C) Values resulting from calculating the area under each of the previous curves. D) Representation of the values obtained after using the data from C) in Equation 1, in the Specific Materials and Methods section. The area under this curve corresponds to the virulence index for that phage in that strain. Error bars represent the Standard Error of the Mean (SEM).

Once these calculations have been made for all the different combinations of phage, strains, and inducers; the resulting values were grouped and represented as columns. Each of the tested phage populations (WT, Fixed Riboswitch, and steps 2, 5 and 16 of the evolutionary procedure) were represented for both positive and negative selections, in the absence and presence of different concentrations of theophylline. These values were then used to calculate the riboswitch's activation efficiency. All these different assessments can be seen in Figure 35, which shows no change whatsoever in the case of WT T7, which was expected, and a general higher activation fold in the case of the Fixed riboswitch sequence. Different phage generations show changes in their virulence, which don't necessarily correlate to an increased activation fold of the riboswitch.

# Recombination-based evolution strategy

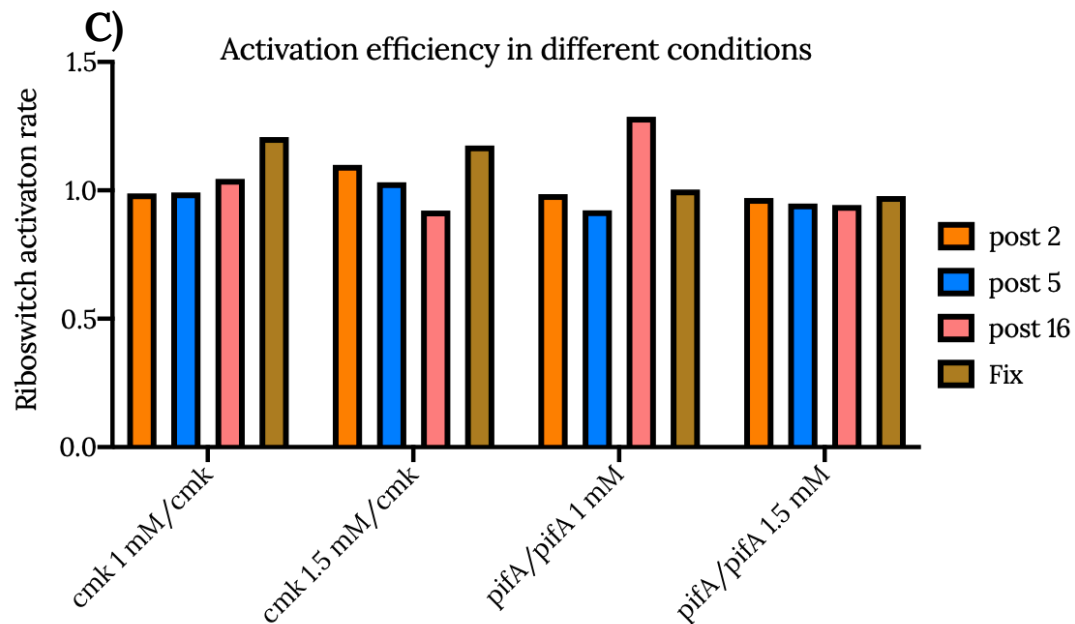
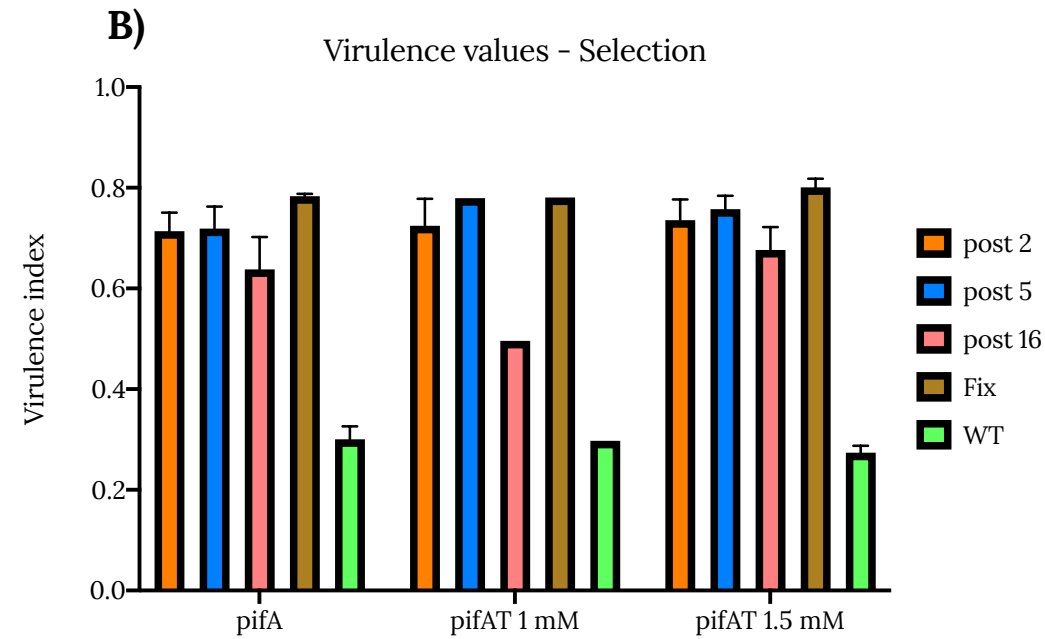
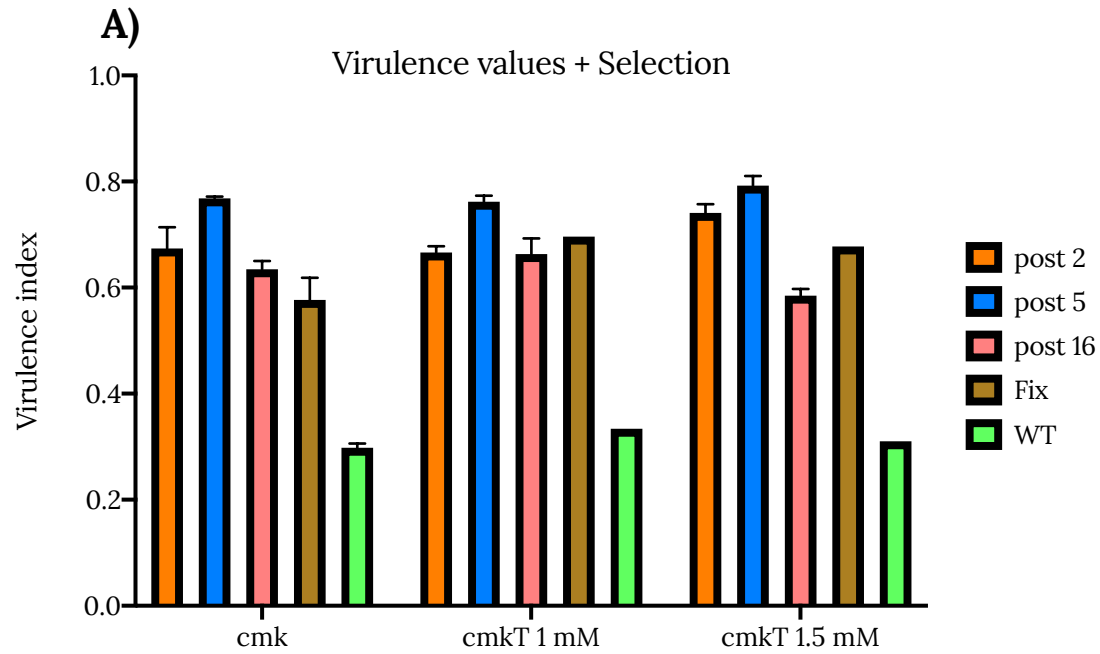


Figure 35. Virulence index comparison between populations.



Virulence index values for the different phage populations (WT, Control/Fix riboswitch, and after 2, 5, or 16 cycles of selection) when infecting the positive and negative selection strains in the presence and absence of different concentrations of theophylline. The higher the bar, the higher the virulence of that particular phage population in that particular strain, with a range between 0 and 1. A) Virulence indexes for the 5 different phage populations after infecting the positive selection strain in the absence and presence of theophylline. In this case, virulence should be higher in the presence of theophylline, as the riboswitch changes to an ON conformation in its presence and allows for the expression of *cmk*. B) Virulence index for the different phage populations when infecting the negative selection strain in the absence and presence of theophylline. In this case, virulence should be higher in the absence of theophylline, as the riboswitch stays in an OFF state. Once theophylline is being added and the switch turned ON, *pifA* is expressed, blocking efficient T7 replication. C) Representation of the fold increase in virulence in both the positive and negative selections, for the two different concentrations of theophylline used in each. In the positive selection the virulence values in theophylline<sup>+</sup> cultures were divided by the value in theophylline<sup>-</sup> cultures, with the opposite being done for the negative selection, as indicated in the figure's legend. Error bars represent the Standard Error of the Mean (SEM).

### 4.4.5 Protein receptor evolution strategy

Based on this same strategy, and in collaboration with Dr. Christophe Corre (University of Warwick), a second evolutionary project was devised; however, instead of riboswitches, the objective in this case was to evolve specific cell receptors working as part of the *mmfR*/*mmyB* system. This set of genes regulates the biosynthesis of an antibiotic molecule known as methylenomycin (Mm) (O'Rourke et al., 2009), with the *mmyB* gene regulating the activation of Mm synthesis, and the *mmfR* gene doing the same for the repressor protein MmfR, that blocks *mmyB* until the autoregulating molecules methylenomycin furane (MMF) reach a high enough concentration. (Figure 21)

The elements involved in the evolutionary processes for this strategy were the same as in the main project, with the difference of the evolutionary target and its regulators.

The project considered two possible evolutionary procedures:

- The evolution of a repressor that would only bind to *mmyB* in the absence of the ligand, but never in its presence, essentially increasing the specificity and orthogonality of the system, much akin to the riboswitch's evolution strategy.
- The second strategy is aimed towards the evolution of mutant homologue versions of the *mmf* repressor *mmfR*, named *sgnR* (Sidda et al., 2014) and *SAV2270* (Corre et al., 2008). Both versions would be presented with *mmyB* and a version developed by the Corre group, coded by a mutant gene labelled *MARE4*, which differs in just a few nucleotides from the original. All done in order to evolve the repressors toward only recognizing the original version of the protein, even when differences are minimal. This was the strategy selected to move forward due to the interests of the collaborating group.

As with the riboswitch evolution procedure, the sequences of interest were introduced inside the phages via homologous recombination and subsequent specific selection. Three different phage populations were obtained, containing the three different versions of the *mmfR* gene previously described. All three versions would undergo the selection procedure, just as described previously, using *cmk* in the positive step and *pifA* in the negative step. However, in this case the dynamics change with respect to the original evolution strategy.

The positive selection plasmid carried *cmk* under the regulation of the *MARE4* mutant, meaning the repressor wouldn't bind in this case, allowing for the expression of the gene, and thusly the replication of the phage. As for the negative selection plasmid, in this case, *pifA* was under the regulation of *MmyB*, meaning that the repressors which worked would bind to it and block *pifA*'s expression, ensuring the correct replication of the phages carrying them.

All four different constructs were assembled for the selection procedures and transformed into *E. coli* BW25113 cells and *E. coli* BW25113  $\Delta cmk$  cells. These cells also contained the plasmid with the original *mmfR* gene, in order to test the

baseline repression efficiency of the system in each condition. Since *mmfR* is regulated by the Lac operon, two sets of experiments were done, with and without IPTG. All these plasmids can be found in Supplementary figure 6 and Supplementary figure 7.

A preliminary version of these results, showing the difference in *mmfR* activation with and without IPTG between different strains as it correlates to GFP expression, is represented on Figure 36.

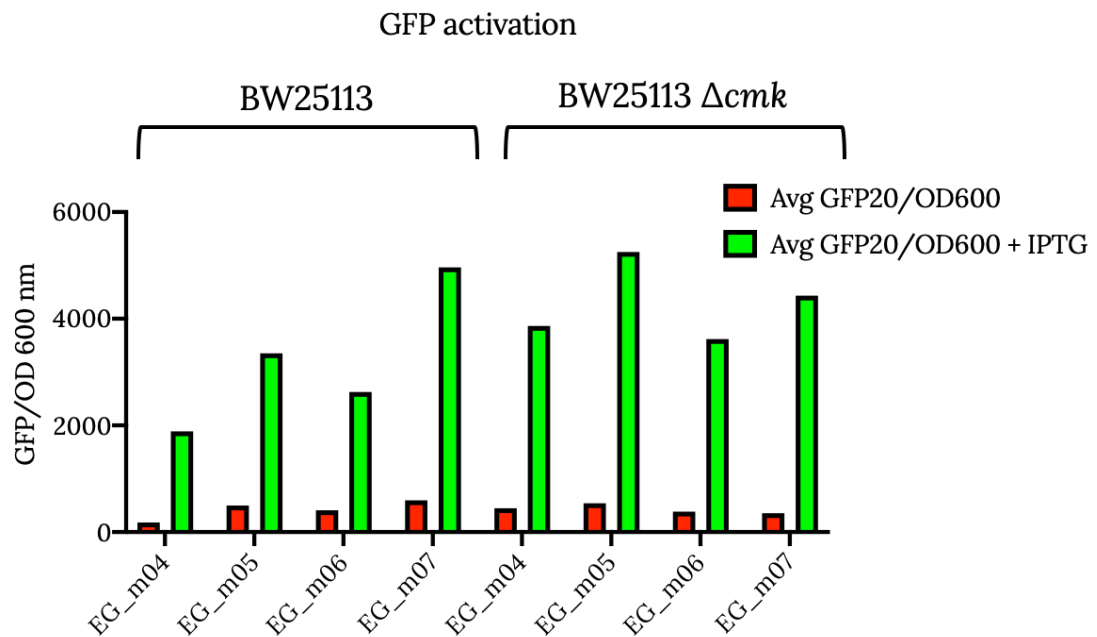


Figure 36. GFP activation in the *mmfR*-*mmvB*/MARE-4 system.

Preliminary results of the GFP/OD values for the *mmfR*-*mmvB*/MARE4 repression systems in *E. coli* BW25113 and *E. coli* BW25113  $\Delta cmk$  cells, in the presence and absence of IPTG. In red, the cases where IPTG is not added into the media, and the Lac operon remains OFF, not producing *mmfR*. In green, the same strains in the presence of IPTG, with an active Lac operon and by extension, *mmfR* production. The values are obtained by determining the steepest point of the growth curve for each strain and dividing the GFP and OD values corresponding to those same timepoints. The selection plasmid present in each case is shown under each set of columns.

These experiments, however, were lacking the presence of T7 phage, which is why another set of tests were conducted to assess their efficiency when infecting and killing the different strains under different conditions; which was done in the form of OD growth curves. Figure 37 shows the results from these assays,

indicating the differences between the control and test strains in the positive and negative selections, with the MARE4 strain showing faster killing due to less binding by the repressor, but negative selection in general showing no major difference at all in most cases. This could be used to confirm the leakage of the system and thus it is possible to use along with the T7 phage in order to produce the desired evolution. Out of the eight strains used, 4 would work as a control, while the others would be used for the selection procedure. All of the investigations were carried out in the presence of IPTG, to activate the Lac operon and thus the *mmfR* production.

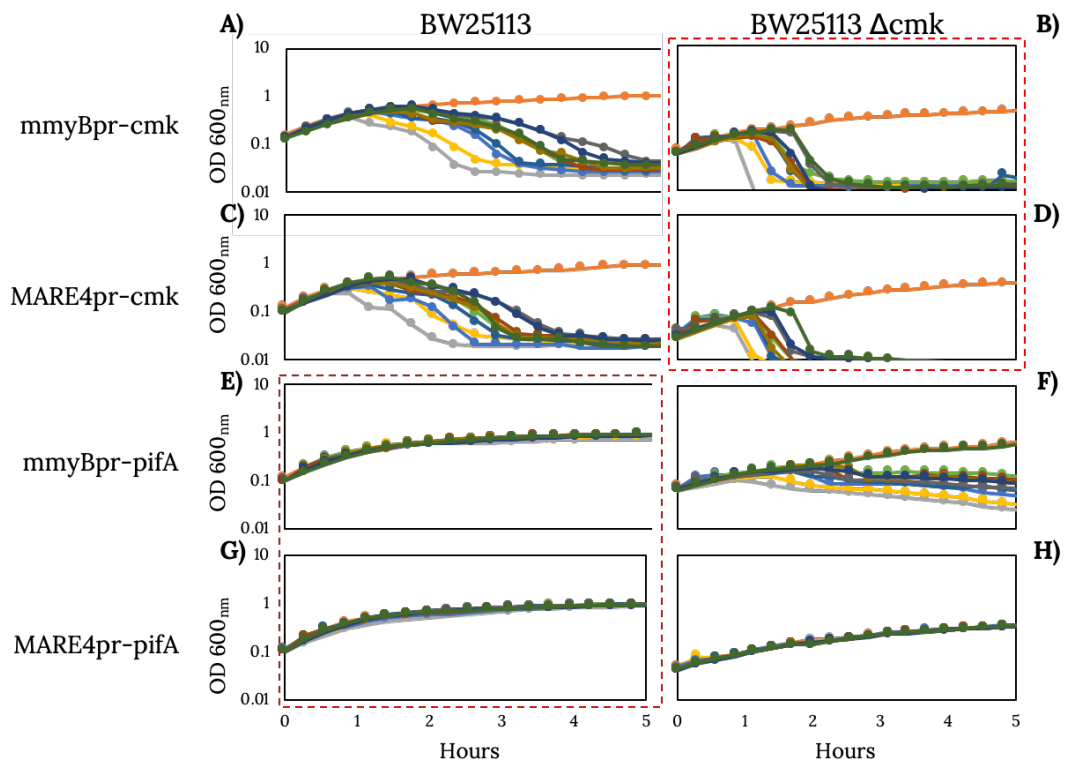


Figure 37. T7 infection of *mmyB*/*MARE4* strains.

Results from the T7 infection of the different strains designed to be used in the *mmfR* evolution

procedure and as controls. The legend indicates the phage dilution corresponding to each element of the plot. The strains used for the actual evolution are surrounded by red dotted lines, while the other ones would be used as controls. In the strains used for positive selection testing (A, B, C, D) different levels of killing are observed. A) and B) show the difference between *E. coli* BW25113  $\Delta cmk$  and *E. coli* BW25113 cells carrying plasmids EG\_m01 and EG\_m04. C) and D) show a similar situation, except that in this case the plasmids carried are EG\_m01 and EG\_m05, so *cmk* repression is lower than in A and B

due to the presence of the MARE4 mutant. E), F), G), and H were used as a way to test regulation in the negative selection step in a similar fashion.

After the testing of the selection strains, the following step was to use homologous recombination to introduce different *mmfR* variants within the genome of the T7 phage, using cells transformed with plasmids EG\_m01, EG\_m02 and EG\_m03. This would allow us to produce recombinant phages carrying variations of the *mmfR* gene, and therefore, evolve three different sequences with the same objective, which would be interesting not just from an experimental point of view, but also as an exercise in possible convergent evolution, structural analysis and gene function.

These phage populations were indeed produced through homologous recombination, but due to several mechanistic issues with the bioreactor, laboratory organization and time constraints of the PhD, the project was not able to move forward and the phages remained unused.

### 4.5 Discussion of results:

Throughout this chapter, the procedure by which a strategy for riboswitch directed evolution was developed and tested have been discussed, including the different ways this evolution was assessed, via the use of sequencing and virulence index assays. Moreover, the adaptation of this process for the evolution of alternative biosensors was also discussed, with some preliminary data showcasing the main elements within this alternative strategy.

#### 4.5.1 Riboswitch evolution

Using a library of recombinant T7 phages carrying randomised riboswitch sequences, the aim of the strategy was to develop a selection system that would allow for the evolution of the riboswitch sequences, which regulated the expression of the  $\lambda$  phage cI repressor. The development and setup of this strategy was arguably the main aim of this PhD, where most time has been spent.

Furthermore, some of the results obtained in this case set up the ground for the development of several other strategies described throughout this thesis.

Results for this strategy can be easily divided into 2 main groups:

The setup of the double selection process, containing a Positive Selection, in which the selected riboswitch sequences are in an “ON” state in the presence of the activator, theophylline; and a Negative Selection, in which they are selected in an “OFF” state in the absence of theophylline.

Once these two selections were tested and properly assembled, several rounds of alternative selections took place, with the objective of evolving the riboswitches. This led to a set of assays used to confirm whether or not Evolution had occurred. This was done via two parallel methods, the analysis of Next Generation Sequencing data, and the calculation of the Virulence factor for phage populations at different moments in the Selection process.

### 4.5.2 Selection process

The setup for the selection process was based on several elements that had already been well established in the literature.

- A working theophylline riboswitch obtained by directed evolution, from Lynch et al., 2007, to act as a control switch to set up the procedure.
- The cI repressor from phage  $\lambda$  (Fehér et al., 2012), and the pRM promoter it binds to (D. Huang et al., 2012) to regulate gene expression
- RBSs of different strengths (Weiss, 2001) to further regulate gene expression.

The results shown in Figure 26 very clearly indicate which of the combinations of pRM promoter and RBS are better than others at allowing the production of *cmk*, but from the beginning there is a very obvious issue, and that is the “leakiness” of the pRM promoter (L. Huang, Yuan, Liu, & Zhou, 2015). Except for the two controls, with either a T7 promoter or no RBS, this is easily seen in all cases, where even in the absence of theophylline to activate the riboswitch (therefore producing cI, which binds to the pRM promoter), there are still phages being

produced, as indicated by the loss of OD600, representing cell killing. At this point, the objective became finding the one combination that gives the highest activation when in presence of theophylline.

This condition eliminated most combinations, save for pRM\_32 and pRM\_33, where some significant differences were seen. This was the reasoning for the experiments shown on Figure 27. In the positive control, we can see that the graph corresponding to the PFU of total phage (phages present in supernatant and inside cells) “jumps” earlier than free phage (phages present in supernatant). That time difference, known as the “Eclipse period” (Adams, 1959; Hyman & Abedon, 2009) roughly corresponds to the time where phages are being assembled inside the cells. Since in the total phage count, cells have been forcefully burst to release the phages present inside them, that explains the reason behind the delay between both cases.

Between both test strains, we can see that there is a significant difference in the number of produced phages when theophylline is present or absent, but due to the fact that the difference was much higher in the case of the pRM\_33 combination, it was decided that it would be the most suitable candidate. In both cases a difference between the total phage and free phage can be seen, but it is merely a difference in number, not so much in time, unlike in the positive control. This could mean that the eclipse period in this case is shorter than the 4 minutes between measurements, so by the time the measurement happens, most phages have already been released into the media, explaining the differences in PFU but the same timeframe.

Strain selection was much more straightforward in the case of the negative selection, since all strains save for one (plus the negative control) produced levels of pifA that were low enough to allow the recombinant phages to kill the cells, as seen in Figure 28. Because of this, initially One-steps assays were not considered necessary in the same way they had been for the positive selection, and the effective growth rate was calculated as further proof that this particular promoter and RBS combination would be the most suitable for the selection.



Despite this apparent simplicity in identifying the ideal strain when compared to the positive selection, when one-step assays were eventually done, it was obvious that the difference between the presence or absence of theophylline was not as large as in the previous step. (Figure 30)

There is an issue with the use of the *pifA* gene that needs to be addressed. Despite it being the ideal gene for an experiment such as the one described, wherein its production will hinder the replication of the T7 phage; the fact of the matter is that the mechanism in which its protein product does so is not known as of this thesis. Pieces of the puzzle are known, such as the necessity of either T7 gene 1.2 or 10 to be expressed in the same cell expressing *pifA* (Schmitt & Molineux, 1991), the fact that the translocation of the phage's genome into the cell is affected (Garcia & Molineux, 1995), or that exclusion happens at the membrane (Cheng et al., 2004). Cells expressing PifA shut down without exploding, a process that could have evolved to avoid a larger-scale infection of the colony.

This leaves enough room for speculation, or in this case, unknown effects that may be influencing the experiment, unbeknown to us. According to the literature, the phage is able to introduce a percentage of its genome without an issue, but after a certain point if *pifA* is expressed, the remainder is aberrant. Some phages manage to enter, possibly even replicate, before the cell shuts down, but they are trapped inside. All these experiments analysed *pifA* as part of the sexual F plasmid. It is possible that the presence of the gene in this environment affects the production of the *pifA* protein, and thusly, the phage exclusion procedure.

That, along with the leakiness of the pRM promoter, could be a possible reason for the differences between the total phage count and the free phage count. In cells with a higher expression of *pifA* due to the activation of the riboswitch, the amount of phages being able to properly enter the cell and replicate before the shutdown is lower than in the ones where the switch is off. However, the leakiness of *pifA* makes such an impact that the difference between the two conditions is much smaller than it should be. By bursting the cells, phages



trapped inside get released into the media, to re-infect again, which would explain the higher numbers in those cases.

Despite this, it was considered that the combination of both steps of selection would be a good test for whether or not the system works as intended.

### 4.5.3 Riboswitch evolution experiments:

The virulence factor calculations provided a comparative measure of the different efficiencies for each of the different phage populations, in a much better way than directly using growth curves.

The WT T7 was used as a control, essentially, since the whole purpose of the process was to measure the activation of the switch and how that affected the virulence. Hence, that explains why in this case the value remains at around 30% efficiency. The other control, the fixed riboswitch-containing phages that were used to set up the procedure, work as expected, with an increase of 20 and 17% for the positive selection, and showing essentially no difference in the negative selection. Given the commented issues with the negative selection, this result, although negative, it's not entirely unexpected.

This riboswitch sequence could then be considered the “golden standard” for the experiment. But at the same time, it can help draw the wrong type of conclusion. The evolution of the phage populations can cause increased virulences, but that does not necessarily mean better riboswitch sequences. It is important to make that distinction, and the reason why the “activation efficiency” metric can give more insight into the actual workings of the riboswitch.

For the actual test populations, two out of three are better at infecting than the golden standard/Fixed riboswitch phage, but that does not necessarily mean better riboswitch sequences, as evidenced by the activation efficiency in both cases. Phage populations with worse riboswitch sequences can have a higher virulence, and it is important to take that into consideration when assessing the efficiency of riboswitch activation based on that data. At this point, it is important

to remember the fact that the all phage populations before generation 10 were selected with 1.5 mM theophylline, while every generation afterwards was selected with 1mM. This is actually visible in the positive selection, as generations 2 and 5 are worse than 16 when infecting at 1mM, while the opposite happens at 1.5 mM.

Regarding the negative selection, as seen before in previous assays, there was a larger variation in the experiments. But the results show the phages that were selected at 1.5 mM have a better activation rate than the gold standard switch, while the one selected at 1mM does worse than all others. But at 1 mM, while the 1.5 mM populations were once again less efficient, population 16 shows the biggest activation rate, with an increase of 28%.

These results seem to indicate there is a difference between subsequent generations, and that even within the selection process, further changes, such as concentrations of the activator, can be considered within the strategy, allowing for more specific riboswitches to be obtained.

To fully assess this, the next generation sequencing of different libraries took place. Based on the initial set of sequences analysed, it seems that indeed each generation has been marching towards a dominant sequence from a sequence space where no particular one was overrepresented, as seen in Figure 33.

There are, however, some not-so-obvious implications for the relationship between riboswitch sequence and specificity that can be drawn from these results, especially given the efficiency of phage population 16 at in cells grown with 1 mM theophylline.

The reduction of theophylline concentration occurred after generation 10, and the results indicate that the tested population works best at that concentration than at a higher one. However, when that is considered along with the data from Supplementary table 2 and Figure 33, we can see that by that point (marked as S12 in the figure), one sequence is already overly represented in the population.

Is it possible that a sequence can change its 3D structure to work better at a specific concentration? We do not have all the necessary data to confirm this hypothesis, but this could be a way of explaining the obtained results. One possible way of addressing this would be to test the virulence of another phage population after generation 10.

### 4.5.4 Conclusion and future improvements

So, in conclusion, these two sets of selection data, when taken together, seem to confirm the following:

1. Directed evolution has indeed taken place using this strategy.
2. Phage virulence does not directly correlate to riboswitch efficiency, as some phages have increased the virulence over subsequent generations, despite carrying less efficient riboswitch sequences.
3. The obtained riboswitches have managed to improve upon the previous generations, and in certain circumstances, a switch ratified by the literature, used in order to set up the system. These improvements include a more efficient functionality at a lower concentration of activator when compared to the initial generations, and a better efficiency rate for the negative selection step.
4. There is room for improvement in the double selection, such as reducing the leakage of the selection genes via the promoter, so the full activation fold of the switch is more discernible.
5. We acknowledge the fact that the mechanism of the PifA protein is still not fully understood. This could mean some effects may be overlooked in the selection and cause a larger variation between results than in the case of the positive selection.

#### 4.5.5 Protein receptor evolution procedure:

This project was developed on the back of the *cl* repressor strategy, the *mmfr* strategy was considered a good way of determining whether a T7-based system could be applied to the evolution of not just RNA switches, but also different protein receptors.

Despite the full strategy not coming to fruition, there are still some relevant sets of data that can be discussed within the frame of this thesis.

As before, the selection genes used were *cmk* and *pifA*, but the whole promoter/repressor system is completely different in levels of complexity. The regulation still occurs at the translational level, with the riboswitch turning ON/OFF, but the evolutionary objective is a protein receptor, which adds levels of regulation that may be affecting the process. So far, we have no data directly indicating if protein production was hindered by the process.

Based on the results shown in Figure 36, the regulation seems to differ between  $\Delta cmk$  cells and regular *E. coli* BW25113 cells as the GFP/OD values are higher on average in the second case. But despite that, in both cases all strains containing the mutated MARE4 version produce higher levels of GFP, which could be a sign of less binding by *mmfR* in those cases, according to our objective.

The second set of experiments, this time in the presence of T7 phage to test its impact in the different strains used for selection, was once again done in all the previously used strains, but only in the presence of IPTG, to activate the *Lac* operon.

Each set of selection strains had four strains, two of them tests and two as controls, as explained in Figure 37. As seen here, since the *Lac* operon is activated in both, *MmfR* is being produced and binding to *mmyB*, thus repressing *Cmk* production, cells in B still provide *cmk* in their genome, allowing for the phages to infect and kill. The results in A show lower killing rates which can be attributed to leakage, as the system is not 100% tight and some *Cmk* is still being produced and allowing proper phage replication. The constitutive lack of *cmk* in C affects killing rates the same way as in A, but lower regulation due to MARE4 presence

allows for faster killing, and the same occurs in D when compared to plot B. (Figure 37)

The negative selection is where the biggest issue arises, as no killing whatsoever occurred in either Figure 37 E, G or H, only in F, which was acting as one of the controls. And even then, the killing rates were much slower than in any of the previous cases. This is possibly due to PifA levels blocking any phage replication in all other cases. This can be interpreted as an excess of laxity from the system, and that the levels of PifA protein being produced even in these cases are more than enough to affect the strains in this fashion. In order to check whether PifA levels have been affected, the ideal thing would be to assess via western blot, which could also be useful to determine the minimum concentration needed for repression of T7 replication to take place.

### 4.5.6 Conclusion and future improvements

In conclusion, this preliminary set of data can tell us the following:

1. The MARE4 mutated variant of gene *mmyB* has a lower binding efficiency to the repressor MmFR.
2. The use of a T7-based evolutionary procedure for this system holds potential but needs to be further regulated.
3. Tests for the positive selection step of the procedure showed promise, with differences in growth curves in the presence of T7 phage; but not so much for the negative selection step, where the system requires further optimization and regulation in order to be usable in this context.
4. Some of that optimization would involve western blot assays to determine the different levels of protein production under each of the conditions, which could help shed a light on the threshold that determines where phage killing starts occurring, especially in the context of the negative selection step.

## CHAPTER 5. TRANSDUCTION-BASED EVOLUTION STRATEGY.

### 5.1 Introduction

Viruses have evolved as an obligatory parasite, and as such require the molecular machinery of other organisms in order to replicate, whether they undergo lytic or lysogenic cycles. However, as with all replication processes, nothing is without flaws. In the same way some viruses can integrate within the genome of their host, in some cases bacteriophages can incorporate DNA fragments from their into their genome. This creates a subpopulation of viral offspring which will then transfer that DNA into whichever new hosts they infect, in a process known as transduction, first described by Zinder & Lederberg (1952)

Traditionally, 3 main types of transduction are considered to exist:

- Generalised transduction
- Specialised transduction
- Lateral transduction

Of these three, only the first two are considered relevant in this case, as they both involve bacteriophages (Griffiths, Miller, Suzuki, Lewontin & Gelbart, 2000).

In the case of generalised transduction, the bacterial DNA introduced in the capsid is essentially random and is just present in a low percentage of the obtained offspring. This can be due to the phage integrating itself in the bacterial genome and taking some bacterial DNA fragment via recombination; or because of some phages' "Headful packaging" mechanism (Alam et al., 2008). Phages that use this mechanism, such as the T4 family (which includes the T7 phage) establish their capsid and proceed to fill it up with genetic material, which, in some cases, can include the host's DNA. The start of this process can be determined by the phage's nuclease detecting a specific *pac* site, or can start randomly (Garneau, Depardieu, Fortier, Bikard, & Monot, 2017)

Once these fragments are inside the virion, these viruses are known as “prophages” or “phagemids”, and will transfer them to their next host, where several things can happen:

- The DNA is broken down by the host.
- The DNA happens to be homologous to the host's and recombination is able to occur.
- The DNA was originally a plasmid and re-circularizes in the host.

In the second type of specialised transduction, specific genetic sequences get included within the phage, depending on the position where the phage integrated itself into the genome of the bacteria. One example of a phage that undergoes this process is the  $\lambda$  phage (Snyder, Peter, Henkin, & Champness, 2013).

This ability of phages to act as a movable element that can be used to deliver specific DNA fragments into a host has been taken advantage of not only in bacteria, but even in mammalian cells, as a way to deliver and facilitate gene therapy (Naldini et al., 1996; Robinson et al., 2003). All this sets transduction as the mechanism responsible for the horizontal transfer of genes, and thus, a very active and relevant player in the process of evolution at the molecular level.

Precisely, the use of a transduction strategy enabled Yosef, Goren, Globus, Molshanski-Mor, & Qimron (2017) to develop a way to increase the host range of the T7 phage by using a specific packaging signal (Chung & Hinkle, 1990b), giving the system a much-needed versatility, and enabling the development of an efficient system for DNA transfer along the way. These experiments served as inspiration for an alternative strategy developed in this thesis, as part of the directed evolution of riboswitches that represents its core.

Despite all its positives, such as high speed and number of phages produced per cycle, there are several points where the original riboswitch evolution strategy via T7 could theoretically be improved:

- Homologous recombination: The process of homologous recombination is, at its core, a rare occurrence, hence why the ideal ratios of phage to cell need

to be calculated. This is then followed by an amplification step of the recombined phages. But the most important issue is that this process could also cause a bias on the phage library when compared to the original plasmid library, decreasing the theoretical number of variants available and effectively reducing the library size.

- **Activation time:** In the full phage strategy, phages carrying the different riboswitch variants infect cells that have been grown in the presence or absence of theophylline, and the riboswitch has approximately 17 minutes to activate and allow the proper replication of its carrier, given this is the approximate time it takes for phages to enter, replicate and burst out of a cell. It could be argued that this time constraint may affect the viability of different versions of the riboswitch during the selection process, giving a wrong estimation of the switches that are able to survive each selection step.

Such were the objectives of this alternative strategy. By keeping the riboswitches in a plasmid library, rather than recombining them into the phages, we could ensure their activation by growing the cells with theophylline in the media for enough time to fully activate any working variants present. And by eliminating the full phage in favour of infective particles/phagemids, we make sure that the replication only happens under very specific terms, and can trigger it when needed, rather than having to accommodate everything in the original seventeen minutes of T7 replication.

The starting T7 phage population in this case is already a KO for DNAPol, and again includes *trxA* as a marker, but not as part of a recombinant sequence of interest as it was before. In this case, the initial infection is carried by these defective phages, and the process occurring is phage transduction (Zinder & Lederberg, 1952). This process can make for higher efficiency than homologous recombination, pack our plasmid of interest inside the phage's virion; and decrease the chances of WT T7 contamination, since there were no WT phages present at any point in the system.



This strategy requires a new, less automatic experimental setup, as well as cells transformed both with plasmid containing the T7 packaging signal, the riboswitch library and the *cl* repressor; and the specific selection plasmid. This tentatively gives the researcher higher control of the selection process as a trade-off, as the riboswitch can be induced for longer than the  $\pm 17$  minutes it takes for the phage to replicate.

Unlike the first strategy, phage replication in this case is regulated from within the bacterial strain, eliminating any issues arising from using WT phages from the beginning, using  $\Delta$ gp5 T7 phages instead. The T7 DNA pol will be supplied by the cell, and the switch's activation will regulate the proper replication of phages, following the same design from the 1<sup>st</sup> strategy. Phages infecting cells containing a working version of the riboswitch will be able to replicate, and a percentage of their offspring will package the plasmid containing it, creating the infective particles or phagemids.

In order to detect these phages and continue the process, a bacterial strain containing the selection plasmid for the next step will be infected with the obtained phages and plated in media containing a specific antibiotic. The gene conferring resistance to this antibiotic is encoded inside the phagemid's plasmid, meaning that unless these cells have been infected by the  $\Delta$ gp5 T7 phages and transduction has occurred, they will not be able to grow.

These cells will then be infected again by a new population of  $\Delta$ gp5 T7 phages, setting the process in motion once again. As indicated before, all this makes the system have lower automation than the initial strategy, sacrificing it for a higher efficiency on the first bottleneck of the process, the size of the initial phage library.

All the elements needed for this strategy were developed and partially tested, but due to time constraints, no double selection procedures were carried out, and thus, its efficiency as a directed evolution method has not been tested.

## 5.2 Specific objectives

To develop and assess a transduction-based procedure founded on the use of the bacteriophage T7. Such procedure would be used for the directed evolution of riboswitches and possibly modified for the evolution of alternative biosensors.

## 5.3 Specific materials and methods

### 5.3.1 Packaging signal design

As previously stated, based on the research by Yosef et al., (2017) and the packaging signal articles by Chung & Hinkle, (1990), several different versions of the packaging signal were designed, and eventually synthesised. Their sequences differed in size, due to the different fragments taken from the short and terminal repeats of the T7 phage's genome. In the end, three different packaging signals were designed, but only 2 were produced. Their general structures and individual differences are shown on Figure 38. In the end, the used sequence contained fragments from both left and right short repeats, as well as terminal repeats.

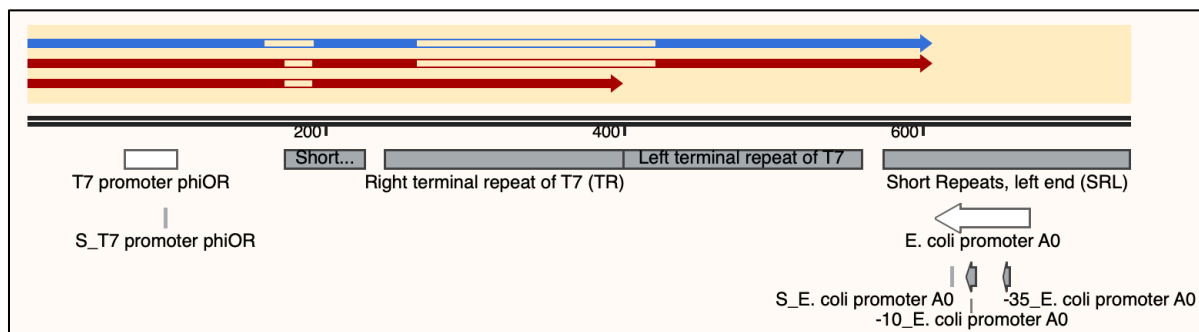


Figure 38. General structure of packaging signals

Structure of the different packaging signals designed for the project. In red, the ones that were actually synthesized, in blue, the one that didn't make it past the design stage. As the figure indicates, all three were very similar to one another when aligned (filled arrow fragments), with some differences regarding the length of certain sequences (empty arrow fragments). The main difference between the two synthesized packaging signals was the inclusion of the left-end short repeats (SRL) at the end of its sequence. The part that shows up as not aligned is essentially the same, given that the terminal repeats on both left and right sides have the same sequences.

### 5.3.2 Plasmid Assembly

Plasmid assembly specifications:

Given that there was only one main modification in order to carry this strategy, and that is the addition of a T7 packaging signal, most constructs for this strategy remain the same as in the case of the full phage one. Selection plasmids remain the same in both cases, and no comparison between the different RBSs is needed in this case, as those were already carried out previously and the same results would be applicable here.

That being the case, there was only one new construct for the *cl* strategy, based on the riboswitch library plasmid EG\_cI02, but its assembly required different steps, and further modifications occurred as time went on and the definitive version was obtained. Following that same idea, new constructs were also designed for the *mmfR* strategy, but were not properly cloned. All the assemblies were made via GoldenGate ligation (Engler et al., 2008), and the final products can be seen in Supplementary figure 8.

- EG\_pm01a (3916 bp, Supplementary figure 8. A):

The initial modification consisted on the elimination of most of gene 4.7, part of the homologous regions), in order to include the packaging signal in its place. As a result of this, this plasmid was obtained. However, further modifications were needed, as both genes 5.3 and *trxA* were still included as part of the sequence of the plasmid.

- EG\_pm01b (5594 bp, Supplementary figure 8. B):

The second step in the assembly, eliminating both the other end of the homology region, gene 5.3, as well as *trxA*, in order to introduce the sequence of the T7 E.P. *gp5*.

- EG\_pm01c. (5582 bp, Supplementary figure 8. C):

The definitive version of the plasmid, eliminating the remaining fragment of gene 4.7, as well as the T7 promoter located in front of the riboswitch, exchanging it for *E. coli* promoter J23119 (C. C. Liu et al., 2012).

### 5.3.3 Production of infective particles/phagemids

An overnight culture of cells transformed with the previously mentioned plasmid was refreshed into 5 ml and grown at 37°C until it reached logarithmic growth phase ( $OD_{600}$  0.2-0.3). At that stage, it was infected with T7 phages and kept at 37°C with shaking until the culture became clear. The cultures were then pelleted by centrifugation at 3000 rpm/2254 g for 10 min in an Eppendorf centrifuge 5920 R. To obtain clean samples, lysates were filtered through 2  $\mu$ m Sartorius filters. 50  $\mu$ l of chloroform were added and the culture was vortexed in order to fully eliminate any possible remaining cell debris. This lysate, however, contained both phages and phagemids, which have incorporated the plasmid and need to be isolated.

### 5.3.4 Transduction assay

In order to confirm the presence of transducing particles in the lysate, the culture to be infected needed to be grown until  $OD_{600}$  0.2-0.3, before 300  $\mu$ l of the culture were mixed with 100  $\mu$ l of the lysate and plated on an LB agar + Cm plate. If transducing particles were present on the lysate and infect the cells, colonies would appear, as the cells would have acquired the plasmid contained within the phages and obtained the necessary antibiotic resistance.

Initial transduction assays were made with cells that only contained the packaging signal plasmid, as the objective was to measure the efficiency of the packaging. The objective was that eventually, the procedure would happen with cells carrying either the positive or negative selection plasmids.

### 5.3.5 Plaque assays

Regular plaque assays took place in parallel with the transduction assays, so as to confirm the presence or absence of full-genome-carrying phages in the lysate. The procedure was as described in the regular materials and methods, with 300  $\mu$ l of the cells that are to be infected + 100  $\mu$ l of the phage lysate. This is then mixed up with 3 ml of soft agar and poured onto a plate with the original resistance carried by the cells. The number of PFUs in the lysate is then compared

to the number of phagemids/CFUs obtained from transduction, and an approximate ratio of efficiency for the packaging signal can be obtained.

### 5.3.6 Colony PCR

When colonies were able to grow on the Cm-containing plates, single-colony PCRs were made so as to assess the presence of the packaging signal plasmid in those cells and then run on a 1% agarose gel. Positive PCRs were purified using the Thermo GeneJET PCR purification kit and sent to Eurofins Genomics for Sanger sequencing in order to confirm the presence of the packaging signal.

### 5.3.7 Strain selection for Evolution procedures

Given that these experiments were already done in the original full phage strategy, and the selection plasmids suffer no modification for this one, it was considered unnecessary to repeat those, and such, the same strains used in one case were applied for the other.

### 5.3.8 Evolution procedure

As seen in the previous chapter, the experimental evolution procedure was originally devised as an automated process, following on the concept of PACE (Hubbard et al., 2015). However, this strategy, following up on (Yosef et al., 2017b) requires extra steps for the transduction due to the use of phagemids, since these cannot be automated using the bioreactor.

Once the packaging signal was proven to work, and following the results obtained from the evolution strategy using recombined phages, a new procedure was derived for the directed evolution of riboswitches using infecting particles/phagemids. In this case, the plasmid carrying the packaging signal is introduced inside the selection strains, which are then grown and eventually infected with  $\Delta gp5$  T7 phages, to avoid any WT contamination. These phages would be able to replicate and pack the riboswitch-containing plasmid thanks to 2 distinct elements present in the cells: the T7 DNA polymerase, which allows for the replication of the phages, and the riboswitch, whose activation will regulate the expression of the specific gene corresponding to the current selection step.

The main difference here for riboswitch function would be the fact that, while in the recombination strategy the activation of the switch and transcription of the downstream genes had to happen in the time it took the phages to infect and burst out of the cell; in this case the switch can be activated for hours before the infection occurs, ensuring the expression or lack thereof of the necessary genes. As explained before, the selection procedure would be regulated by the same genes.

In the positive selection, where an “ON” state of the switch is selected, the essential gene *cmk* would be produced, allowing for the replication of the phage and packaging of the plasmid in those cells containing a working riboswitch sequence in the presence of theophylline.

In the next step, the negative selection, where the “OFF” state is selected, T7 exclusion gene *pifA* is produced in the cases where riboswitch sequences still manage to remain in an “ON” state despite the absence of theophylline. This way, only those containing an “OFF” state switch will be able to replicate and package properly.

Lysates obtained from each of these selection steps would then be used for transduction assays using the next selection strain, in a reiterative manner that would allow for the selection of riboswitches over time. In this step, the phagemid particles that have packed the riboswitch-containing plasmid are used to infect the next selection strain and transduce those plasmids into the cells. Doing so confers these cells an antibiotic resistance (chloramphenicol) and allows for their selection on plate and subsequent growth.

A schematic diagram of how this procedure would be carried out, with an overview of the main elements and steps involved is shown in Figure 39.

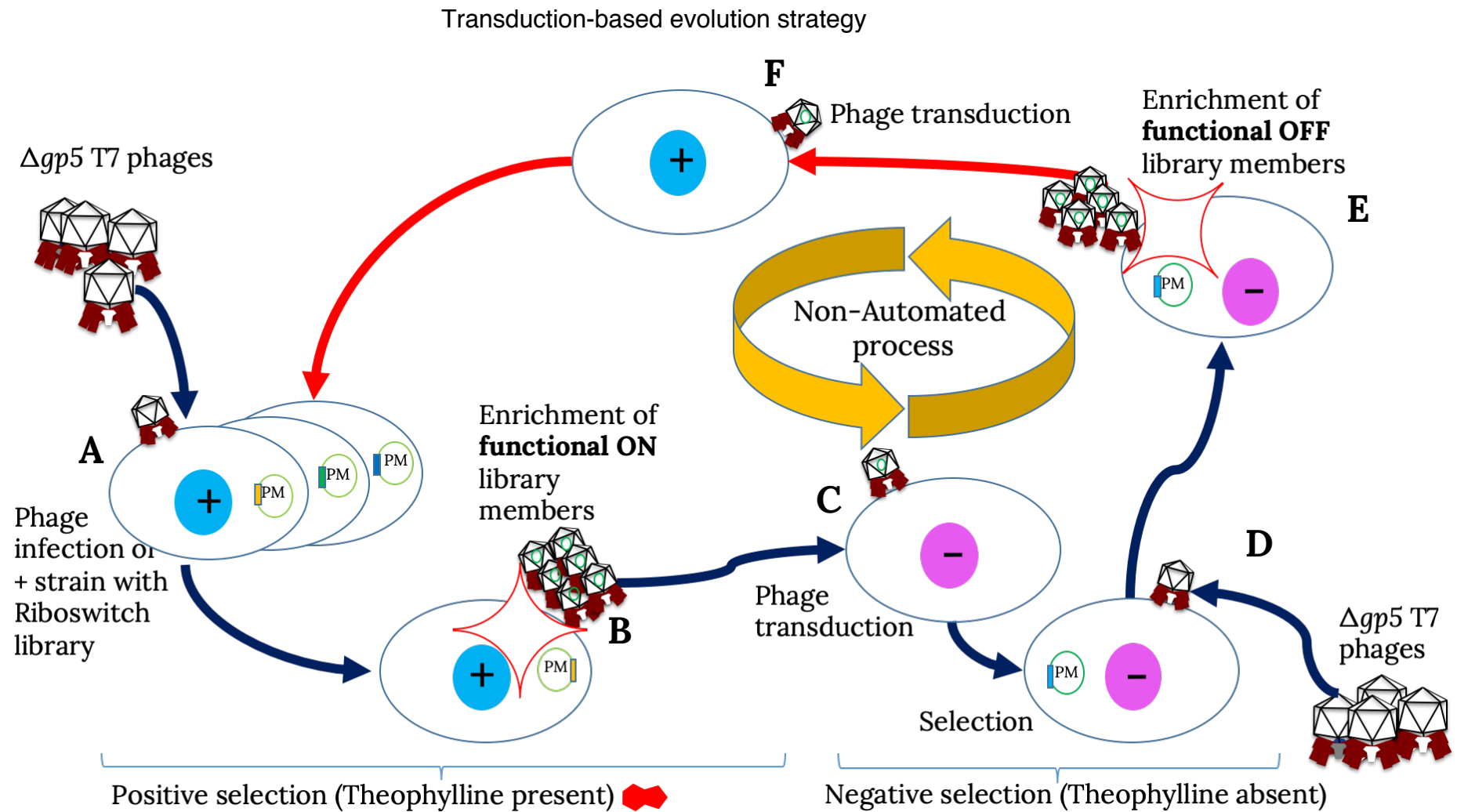


Figure 39. General structure of the Phagemid evolutionary procedure.

Schematic representation of the different steps needed as part of the planned evolutionary procedure using infecting particles. As indicated, the  $\Delta gp5$   $trxA^+$  phages would infect an *E. coli* BW25113  $\Delta cmk\Delta trxA$  strain (A), containing both a positive selection plasmid (blue circle), and a riboswitch library plasmid with the packaging signal (PM), all this in the presence of theophylline, to turn the riboswitch into its “ON” conformation. Phages infecting cells containing functional riboswitches will then be able to activate the expression of *cmk*, and thus allow for phages to replicate/package properly (B). The obtained lysate will then be used to infect *E. coli* BW25113  $\Delta trxA$  cells carrying the plasmid for the negative selection and then plated (C). Cells incorporating the plasmid will be able to grow in cm plates, then grown and be infected once again with the same phage strain (D). In this case, due to the absence of theophylline, the selection is aimed towards the “OFF” state of the riboswitch. If any riboswitch is in an “ON” state, T7 exclusion gene *pifA* would then be expressed and phage replication will be hindered (E). The corresponding lysate will then be used to infect the positive selection strain (F), restarting the cycle.

## 5.4 Results

### 5.4.A Packaging signal assessment

The initial set of experiments carried out for this alternative strategy were essentially aimed at determining the efficiency of the packaging signal sequence included within the plasmid. As a way of determining it WT T7 was used as an initial test, and if results were positive, eventually other phage populations would be tested for the transduction procedure.

- WT T7, as previously said, acted as an initial test and control for the process, since having a full genome meant any problem in the process could solely be blamed on the packaging signal.

Eventually, some of the other phage populations used included:

- T7  $\Delta gp5$   $trxA$ , to determine if the presence of *gp5* in *trans* affected the process, and because this was intended as the phage strain that would be used once the evolution procedures were started. It was used to infect *E. coli* BW25113 cells.



The packaging signal plasmid was transformed into two different bacterial strains, conferring them chloramphenicol resistance. These cells were then used for the production of transducing particles, as explained in the materials and methods.

- *E. coli* BW25113, as a control with all essential genes needed by the phage. In the case of infection with WT, ideally there should be production of both “full” phages carrying the whole genome, and phagemids, carrying the plasmid. In the case of  $\Delta gp5$  phages, since the plasmid carries a copy of the gene, there may be differences, but both phages and phagemids should also be produced.
- *E. coli* BW25113  $\Delta cmk$ , to assess the impact of the absence of *cmk* in packaging. This is a very relevant test, as *cmk* is one of the elements used in the selection procedure, specifically in the positive selection; and ideally should affect all phage strains. Ideally, no phage nor phagemid production should occur in this case, due to the essential role of the gene, which allows the phage to break down the cell’s chromosome in order to be able to use its nucleotides for its own replication (Qimron et al., 2006).

After growing for 1 hour in the presence of the phage, the different cultures were pelleted and filtered, and the obtained lysates then used for their own sets of transduction and plaque assays.

For each lysate, the assays were carried out with several strains of *E. coli*, to determine the effect certain genes had on phage transduction and replication. Each of the selected strains were chosen for their differences in allowing proper phage replication. Being tested in that case, the same strains were also tested for possible effects on transduction, although ideally, no significant differences should occur in that aspect, as the process merely involves the introduction of the genetic material carried within the phage into the cell.

*E. coli* BW25113, as a control with all essential genes needed by the phage.

- Plaques: In this case only plaque assays with WT phage lysate should show any clear spots, as the other populations lack the presence of T7 DNA polymerase.
- Transduction: There should be no impact, as the phage merely introduces the plasmid inside the infected bacteria.

*E. coli* BW25113  $\Delta cmk$ , to assess the impact of the absence of *cmk* in transduction.

- Plaques: Ideally, no plaques should appear with any of the lysates.
- Transduction: possible colony presence may be affected due to the gene product's function, as it interferes with the cell's chromosome

*E. coli* BW25113  $\Delta trxA$ , testing whether the absence of *trxA* affects the process.

- Plaques: The impact should be felt in all three phage strains, as WT lacks the essential gene, while the other two carry it in their genome, but don't have T7 *gp5*/DNA polymerase.
- Transduction: Not much effect would be expected, as the gene product does not affect the host directly.

*E. coli* BW25113  $\Delta trxA$  *gp5*, essentially to confirm whether phages carrying *trxA* in their genome can function if there is a supplied copy of T7 DNA polymerase.

- Plaques: Only WT plaque formation should be affected in this case.
- Transduction: Colony formation results should be similar to the previous case.

TOP10: To confirm if there is a difference between the efficiency of the process between *E. coli* strains, as the genotype varies between them, with TOP10 being *rec<sup>-</sup>*, which may have an impact on the transduction efficiency. Previous experience in the lab has shown TOP10 to be less susceptible to T7 infection.

- Plaques: In theory, only the lysate obtained from WT phages should produce plaques, the same way as with *E. coli* BW25113, but accounting for the possibility of them being lower in number.
- Transduction: No impact, but as with plaques, there is the possibility that the number would be lower than with *E. coli* BW25113.

The results from the initial packaging signal efficiency assays are shown on Figure 40, where the number of Colony forming units, which correlates to the number of infective phage particles, is shown after infection for each *E. coli* strain and lysate. It shows that the number of CFUs produced after transduction varies between the different strains infected with the lysates, but most importantly, that the production efficiency of infective particles is highly affected in *E. coli* BW25113  $\Delta cmk$  cells, as barely any CFUs were produced when that lysate was used, indicating a lower number of them present in the lysate.

### CFU numbers in different strains post transduction

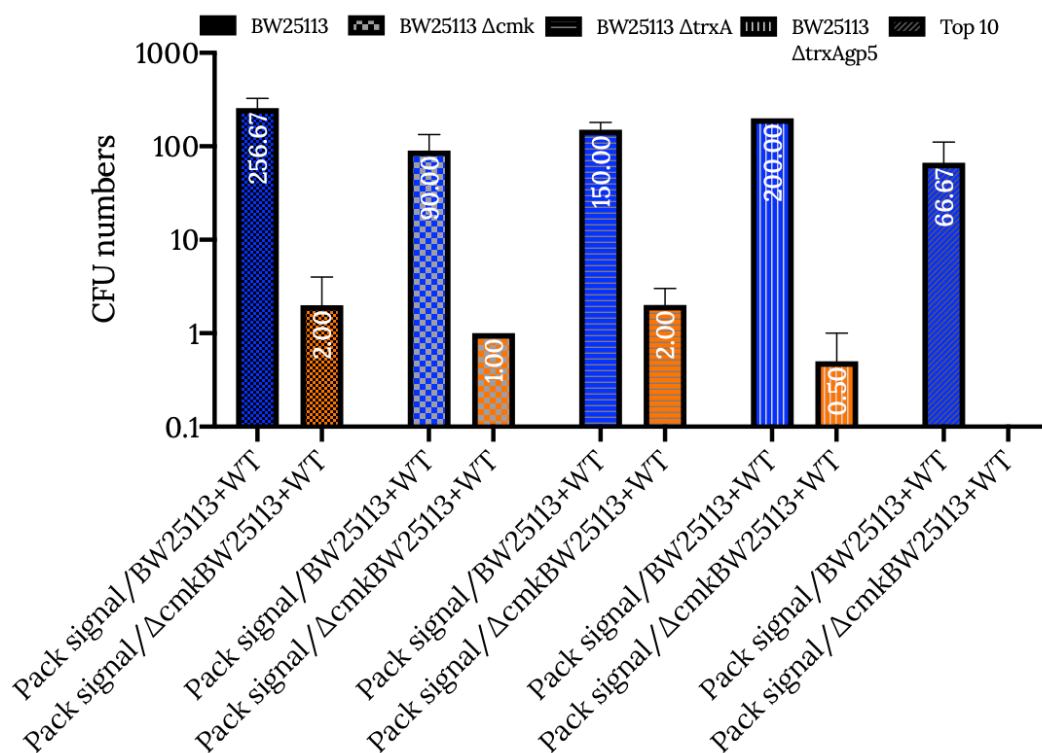


Figure 40. Number of colonies after transduction assays.

Preliminary results from the transduction assays using lysates produced after the infection of WT T7 phages into different *E. coli* strains containing the packaging signal plasmid. The blue bars, corresponding to the infection of *E. coli* BW25113 cells with the WT phage, shows the highest number of colonies in all the different transduction cases. The orange bars represent the cases where the initial infection took place in a  $\Delta cmk$  cell strain, showing a very significant decrease in the number of colonies. The values represented were obtained from 3 independent replicates and the Y axis is represented on a log 10 scale. Error bars represent the Standard Error of the Mean (SEM).

Figure 41 shows the number of Plaque forming units/Phages produced after infection for each *E. coli* strain and lysate, and how viable phage production varies in each of the infected strains, but also, that essentially no phages were produced in lysates obtained from *E. coli* BW25113  $\Delta$ cmk cells.

### PFU numbers after infecting different strains

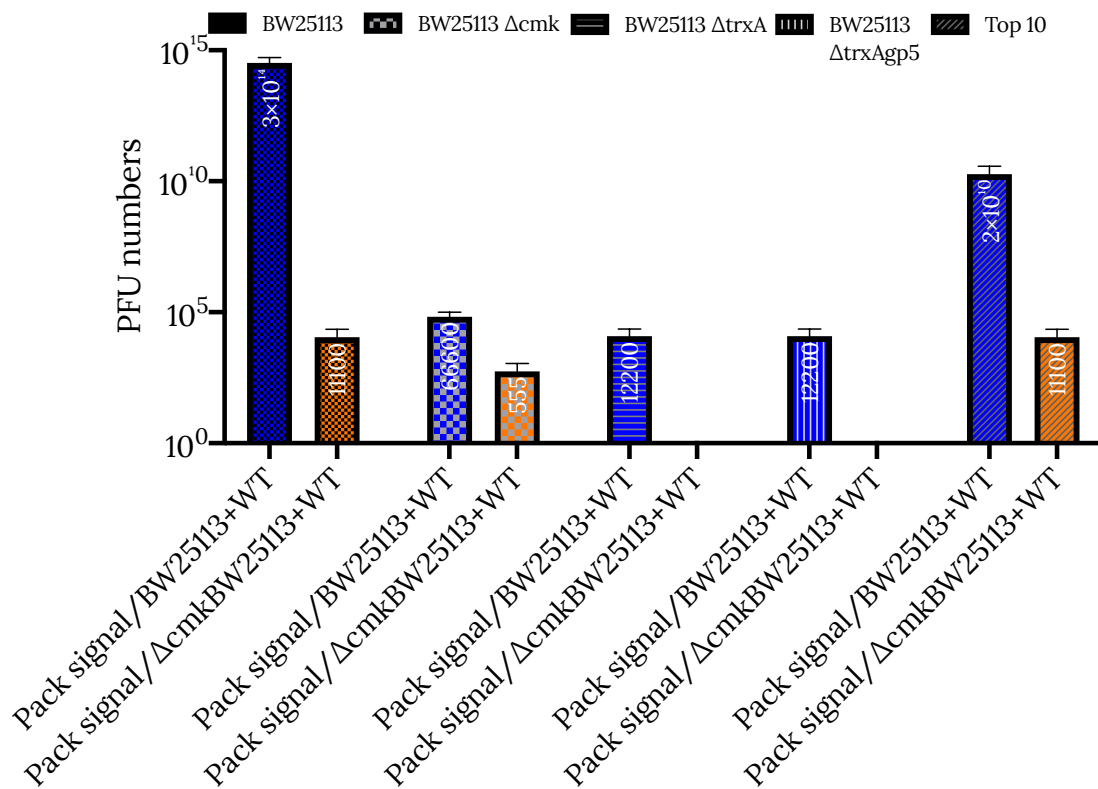


Figure 41. Number of PFU after plaque assays with alternative lysates.

Preliminary PFU values for the different lysates obtained after the infection using WT T7 into different *E. coli* strains containing the packaging signal plasmid. The blue bars represent the lysate obtained from infecting *E. coli* BW25113 cells containing the plasmid, while smaller or non-visible orange columns represent the PFU number present in lysates obtained from *E. coli* BW25113  $\Delta$ cmk cells. The most accurate value for the PFU content corresponds to the one obtained when using the lysates to infect *E. coli* BW25113 cells, as there is no impediment present that could hinder the proper replication of the T7 phage. The values represented were obtained from 3 independent replicates, and the Y axis is represented on a log 10 scale. Error bars represent the Standard Error of the Mean (SEM).

Given these results, and as previously explained, a second set of assays took place to confirm the transduction efficiency of  $\Delta gp5\ trxA^+$  phages. These results can be seen in Figure 42, indicating the number of colonies produced after transduction assays using lysates obtained from the infection of *E. coli* BW25113 by either WT or  $\Delta gp5\ trxA^+$  T7 phages. In all cases, it seems the phagemid production is impaired in the second case but would still be usable in the desired strategy.

### CFU numbers in different strains after transduction

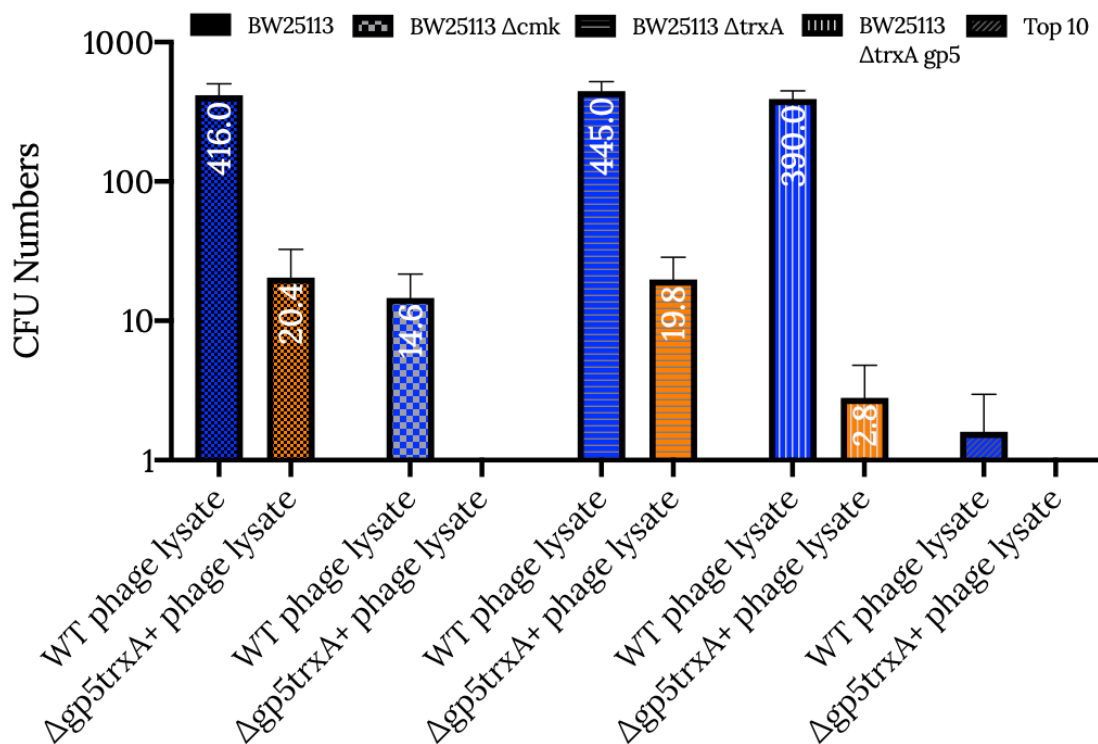


Figure 42. Number of CFU after transduction assays with alternative lysates.

Initial results from the transduction assays using the lysates obtained from the infection of *E. coli* BW25113 cells with either WT T7 phage (blue) or a  $\Delta gp5\ trxA^+$  T7 phage population (orange). The numbers show a lower efficiency in this case than with WT phages, but still colonies are formed, which indicate these phages can be used for the evolutionary procedure once set up. These results were obtained after 5 replicates, and the error bars represent the Standard Error of the Mean. Error bars represent the Standard Error of the Mean (SEM).

## 5.5 Discussion of results

### 5.5.1 Packaging signal efficiency

Based on the results displayed in Figure 40 and Figure 41, the first conclusion that can be drawn is that infecting particle production is in no way close to phage production. This poses a problem for the strategy, but one that can be dealt with by improving the efficiency of the system. There is always the possibility of developing a new packaging signal and a new set of assays to test it, and improving the transduction plasmid, making it smaller and easier to include within the T7's virion. Given the results from Figure 42, the first step in that direction could be the elimination of gp5, as it is also present within the selection strains and its absence doesn't seem to affect packaging. A set of transduction assays with that new plasmid to confirm whether or not phagemid production is impaired would be a good test for future developments of this strategy.

Another plausible conclusion that can be taken out of these results is the fact that packaging is indeed affected by the absence of genes essential to the phage, such as *cmk*. While in the case of regular *E. coli* BW25113 cells the numbers of both PFU and CFU vary between the different cells infected by the lysate, there are enough to formally assess their presence and confirm that the infection of cells containing the packaging signal plasmid is working to a certain degree. However, the same cannot be said when the lysate used has been obtained from a  $\Delta cmk$  cell strain. *cmk* encodes for the dCMP kinase, responsible for the conversion of the host's nucleotides into nucleosides that would be used to assemble the phage's DNA, as a substrate for the T7 DNA polymerase (Qimron et al., 2006).

This can explain the differences seen when compared with regular *E. coli* BW25113 cells. The literature already showed that KO cells for *cmk* grew slower than regular ones, but most importantly, did not allow for the replication of phages, since the phages were not able to properly utilise the host's genome to replicate themselves.

This is one of the issues that make *cmk* such a good element for the positive selection in the evolution of the riboswitch. Unless it is present, phages won't be

able to replicate, and it should only be present when the riboswitch is able to shift to an “ON” state in the presence of its activator, theophylline.

Now, this still does not deal with the inefficiency of the packaging signal when compared to regular phage replication. Despite this difference, however, produced colonies indicate the selected packaging signal was working, and producing enough number of colonies so that, in theory, the evolutionary strategy could be carried forward.

### 5.5.2 Conclusion and future research:

Due to several reasons, chief among them the time constraints of a PhD and the development of the main evolutionary strategy, there is still quite a significant room for improvement in the case of this project.

Suggested improvements:

1. The development of another packaging signal sequence, to increase the phagemid-to-phage ratio of the original packaging signal, thus increasing the overall efficiency of the evolutionary strategy.
2. Fully test the selection procedure. This is the most relevant element of the strategy, and ironically, the one that was not able to be tested properly. This could be done through the transformation of the positive and negative selection strains with EG\_pm01c. These cells would then be grown in liquid culture in the presence and absence of theophylline, and infected with the  $\Delta gp5\ trxA^+$  phages. Since the riboswitch is present in the cells under an *E. coli* promoter in this case, there should be more than enough time for those cells carrying a working variant of the riboswitch in the positive selection and a constitutively “ON” riboswitch in the negative to produce the corresponding selection genes, and support/hinder T7 replication. In the end, if the process is successful, there should be a visible difference between the cultures grown in the presence or absence of theophylline in both cases.

## CHAPTER 6. Q $\beta$ PHAGE INDUCIBLE RNA SYSTEM

### 6.1 Introduction:

The discovery of the CRISPR system has ushered a new era in molecular biology, challenging what was thought to be feasible until now in genetic modification and allowing for very specific genetic changes in multiple different settings (F. Jiang & Doudna, 2017; Makarova et al., 2011). Its understanding has also shed some light onto some of the ways bacteria respond to the many mechanisms by which external DNA is forcefully introduced (Hille et al., 2018). These responses can make certain types of research cumbersome and frustrating, or even eliminate the possibility of some strategies, as the genetic material can get degraded once inside the cells (W. Jiang et al., 2013; Marraffini & Sontheimer, 2008). It is because of these reasons that the development of an RNA-based system that could potentially bypass CRISPR would be interesting, moreover, one which could be relevant for many research groups at different levels of bacterial genetic research. It could even be an alternative on a therapeutic level, eliminating the obstacles CRISPR-Cas systems pose in these particular contexts (Galizi & Jaramillo, 2019; You et al., 2019).

To develop this such a system, the best starting point would be the organisms that evolved to be able to introduce their genetic material inside of bacteria and overcome their defence mechanisms: bacteriophages. Not just any bacteriophage would work in this situation, they need to have an RNA genome for the strategy to work.

The enterobacteria phage Q $\beta$  is a phage of the Leviviridae family, and one of the smallest known phages (Gorzelnik et al., 2016), with a capsid diameter of 28 nm, and a small linear ssRNA genome consisting of 3 Open Reading Frames (ORFs) coding for 4 proteins: a maturation protein, major and minor capsid proteins (contained within the same sequence, but with a frameshift event, in a great exercise of genome economy), and a replicase (D. Brown & Gold, 1996; Kashiwagi & Yomo, 2011). Due to the (+) ssRNA structure, the genome of these phages is not affected by the CRISPR system, which as far as it is known, shears dsDNA



sequences in order to disrupt them. This is what makes the Q $\beta$  phage an interesting and novel vector for the insertion of genetic material into bacteria in a semi-permanent way. Using the phage's strategy to introduce exogenous genetic material into the bacteria, we hope to obtain a vector that would allow for inducible, CRISPR-resistant gene expression, even after the elimination of the original vector and the inducing compound. A schematic version of this process and its main steps can be seen in Figure 43.

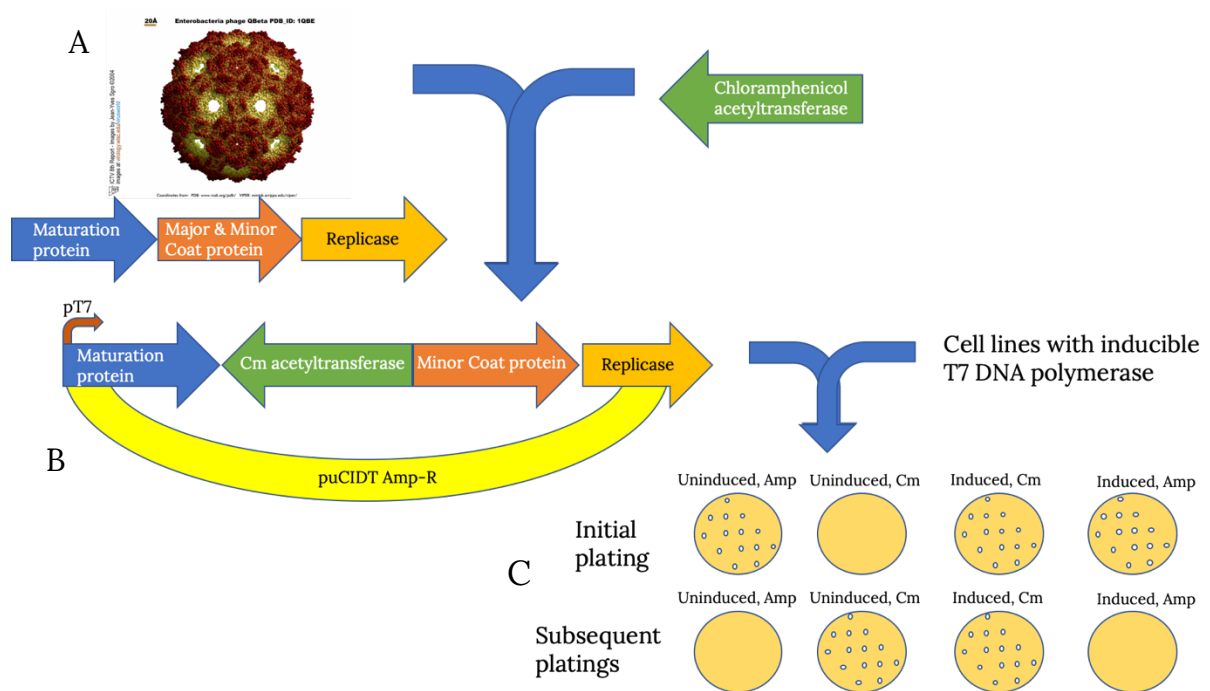


Figure 43. Q $\beta$  phage-based strategy overview.

Schematic representation of the experimental process. A) External and genomic structure of the Q $\beta$  phage, indicating the three ORFs it carries in its genome, encoding for the Major and Minor maturation protein, Coat protein, and Replicase. B) The Chloramphenicol (Cm) Acetyltransferase gene is introduced inside it, instead of the gene for the Major Coat protein, in a reverse sense. This is then contained within a commercial vector carrying Ampicillin (Amp) resistance and transformed into specific cells carrying an inducible T7 RNA polymerase gene. C) Once plated, cells were grown on Amp plates to be selected, and then induced. Induced cells would grow both on Amp and Cm plates; but over the course of generations of selection, they would gradually lose the ability to grow in Amp plates, due to the loss of the original vector, but retain the Cm resistance.

To test this hypothesis, we designed and synthesised a fragment consisting of the full genomic sequence of the Q $\beta$  phage, minus the major capsid protein. In its

place, we incorporated the sequence corresponding to the Chloramphenicol acetyltransferase gene from *Staphylococcus aureus* (Shaw & Brodsky, 1968) in a 3'-5' sense, opposed to the rest of the phage's genome. The entire fragment was placed downstream from a T7 promoter, inside a generic IDT pUC plasmid carrying an ampicillin resistance; and transformed into *E. coli* strains containing an inducible promoter controlling of the expression of the T7 RNA polymerase gene. A schematic map of the construct can be seen in Supplementary figure 9.

The transformation of such a construct conferred one antibiotic resistance to bacteria prior to induction (Amp), and a second one afterwards (Cm). Moreover, this resistance was conserved throughout subsequent generations of cells without further need of induction. After multiple cell passages solely in the presence of Cm, the original ampicillin resistance seemed to disappear, while the induced one remained. But this resistance was also eventually eliminated after several passages with no antibiotic present, proving it was not a fixed condition.

## 6.2 Specific Objectives:

To develop and test a procedure for the inducible and transient expression of genetic material in bacteria based upon the use of the Q $\beta$  bacteriophage.

## 6.3 Specific materials and methods

### 6.3.1 Cell strains

Three cell strains were initially used for this procedure, all of which contained an inducible T7 RNA polymerase. This induction was regulated by IPTG in the case of strains BL21 DE3 (NEB #C2527) and LEMO DE3 (NEB #C2528H), and by arabinose in BL21 AI (Invitrogen #C607003).

### 6.3.2 Constructs specific to the Q $\beta$ project

Maps of all constructs can be found in the Supplementary Materials section.

- pUC-*cat* plasmid: 6767 bp, Supplementary figure 9.

A standard pUC plasmid from IDT carrying an ampicillin resistance was used as a carrier for the sequence of interest, containing the Q $\beta$  phage's replicase, minor coat protein, and maturation protein on a 5' to 3' sense, as well as the sequence for the Chloramphenicol acetyltransferase from *S. aureus* in the opposite direction. Specific primers were designed in order to amplify fragments of the plasmid, so as to check its presence on the cells after transformation, as well as along the curation process. These primer sequences can be found in the corresponding table in the General Materials and Methods Section.

- pUC-*gfp*: 6851 bp, Supplementary figure 10.

As a second stage of the research, a different construct was assembled based on the pUC-*cat* backbone, which contained all the different elements of the previously described plasmid, but the *cat* gene was exchanged for *gfp* in the same orientation. This was done via PCR amplification of the backbone and the *gfp* gene with the addition of appropriate restriction enzyme sites. After the restriction digest, both fragments were ligated and transformed into BL21 AI cells, which were then grown on LB plates (Amp). The presence of *gfp* was confirmed via Sanger sequencing using primers EG\_128 to EG\_133, as seen in the corresponding table in the General Materials and Methods Section.

- Gene fragment T7-rGFP-rRBS-3'UTR: 980 bp, Supplementary figure 11 .

This gene fragment was developed to enable testing in parallel in the PURE system with the full plasmid construct, to confirm the production of GFP in a cell-free environment. It contained a T7 promoter in 5'-3' sense upstream of the sequence for *gfp*, RBS B0034 (Weiss, 2001), and a 3'-UTR. This fragment will not produce any protein alone, but if dsDNA from pT7-RBS-rep is added to the mix, it will produce the replicase, which will bind to the rev3'UTR\_MDV1 and lead to the production of GFP.

### 6.3.3 Primers

Multiple primers were designed for the assembly of the different constructs, as well as the confirmation of the presence in the cell via colony PCR. All can be found in Table 5, in the General Materials and Methods section.

### 6.3.4 Induction

The cell cultures were grown at 37° C until reaching an OD of roughly 0.4, when inducers were added at a concentration of 400  $\mu$ M in the case of IPTG, and 0.1% w/v arabinose in 5 ml cultures. The cultures were induced for 2 hours and then plated in 20 ml LB agar plates with either Amp or Cm.

### 6.3.5 Curation procedure

In order to force the elimination of the plasmid's antibiotic resistance, the procedure was based on growing cells in media without selective pressure, where the genes to be cured were not necessary for the cell's survival, and over several generations, would be eliminated. In the first set of experiments, induced and uninduced cells were plated on LB agar plates with Amp or Cm.

From the colonies that were able to grow on Cm plates, single colonies were picked and grown again in 5 ml of LB Cm and plated once more. The repetition of this procedure over time lead to the eventual disappearance of the original ampicillin resistance carried on the plasmid; while at the same time leading to the conservation of chloramphenicol resistance carried within the genome of the Q $\beta$  phage.

Once chloramphenicol resistant strains had been obtained, a similar procedure took place, with the difference that in this case, cells were grown in the absence of antibiotic, plated in both chloramphenicol and no antibiotic plates, and grown once more from the latter ones. This led to the disappearance of the chloramphenicol resistance, obtaining once again cells that didn't carry any resistance.

## 6.4 Experimental procedure and Results:

### 6.4.1 Q $\beta$ -Cat experiments

After the correct assembly of the plasmids containing the Q $\beta$  genome along with the CAT sequence, the three different strains were chemically transformed, according to the procedure described in materials and methods. After the transformation and plating of the different transformed strains in ampicillin-containing plates, several colonies were able to grow, hinting at the presence of the plasmid. Diagnostic PCR on those colonies using specific primers amplifying the plasmid backbone, confirmed the presence of the plasmid in some of them, as seen in Figure 44.

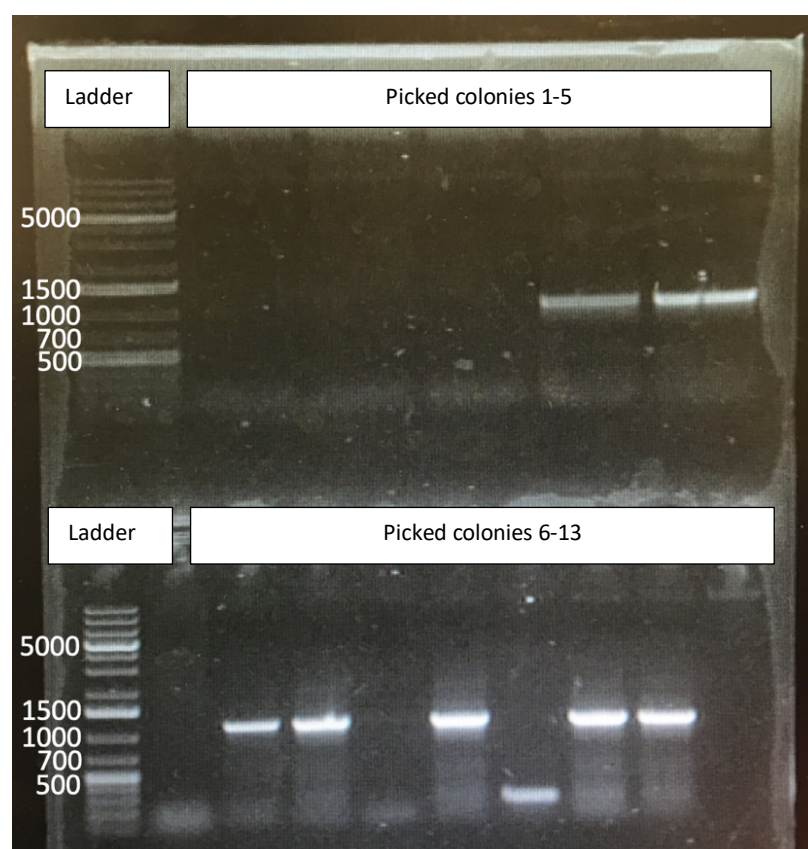


Figure 44 . Diagnostic PCR for Q $\beta$  presence.

Results from the PCR amplification of several of the plated colonies post-transformation of the plasmid construct, using specific primers designed to amplify a fragment of 1157 bp from the backbone. When compared to the sizes of the different bands present in the used Ladder, we could confirm the presence of the desired band, and thus, the presence of the plasmid in those colonies. The ladder used corresponded to the GeneRuler 1kb Plus.

Out of the colonies shown to contain the plasmid, one was selected for each strain, grown overnight and then refreshed into separate liquid cultures both with and without their corresponding inducers. Following that, all cultures were plated in both Amp-LB agar plates and Cm-LB agar plates.

The promoters present in these strains were induced by either IPTG in the case of BL21 DE3 (Briand et al., 2016), or by arabinose in the case of BL21 AI (Guzman, Belin, Carson, & Beckwith, 1995). A third strain, LEMO DE3, was also used, with the difference that although it was induced by IPTG, it had tuneable levels of expression via the use of L-rhamnose, which regulates the expression of the *lysY* gene, coding for a specific T7 DNA pol inhibitor. The induction procedure is discussed in the material and methods section.

From this initial set of plates, only BL21 AI was able to grow on ampicillin in both cases, but not on chloramphenicol unless induction had happened. LEMO DE3 showed growth on both ampicillin and chloramphenicol despite not being induced, while BL21 DE3 showed no growth in chloramphenicol even when induced. After several repetitions, including the addition of various concentrations of L-Rhamnose to the LEMO DE3 strains; this was indeed confirmed to be the case, meaning that only one of the three strains (BL21 AI) was actually inducible enough to be used for this type of experiments. A diagram of these results is represented in Figure 47. These results eliminated these strains from further use in the strategy, meaning that only BL21 AI was used from this point onwards, as it showed growth in chloramphenicol only after induction.

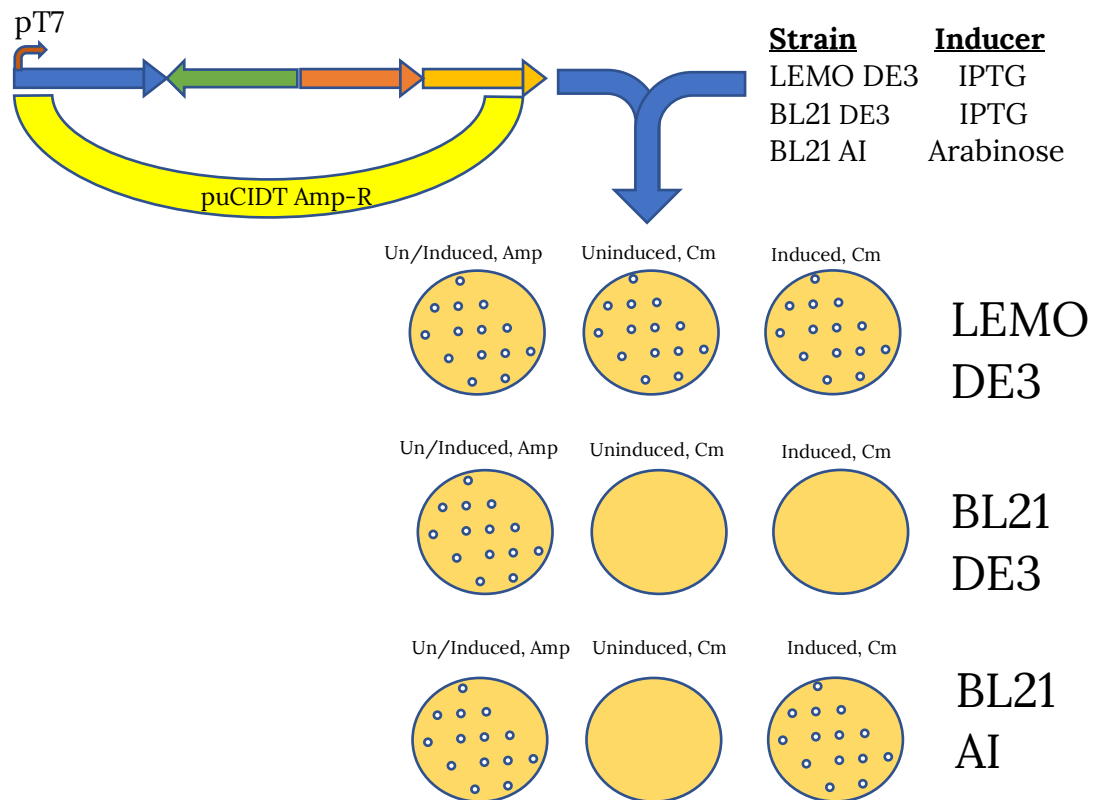


Figure 45. Induction results.

Representation of the results after the induction of different strains transformed with the Q $\beta$  plasmid, indicating whether or not the cells were able to grow in different conditions.

From the BL21 AI cells, those induced colonies which were able to grow on chloramphenicol plates were then grown overnight into 5 ml of LB with chloramphenicol, and plated again on chloramphenicol plates, showing the cell's ability to grow in the presence of the antibiotic even without undergoing induction a second time; something that was not initially considered.

After several iterations of this process of culture and plating, the growth of the liquid cultures and plates containing ampicillin started to show differences with the ones containing chloramphenicol. One possibility for this behaviour could be the cells "curing" themselves of the transformed plasmid, due to the ampicillin resistance gene not being needed any more, but the chloramphenicol resistance remaining within the genome of the Q $\beta$  phage, in the form of RNA.



After 10 rounds of plating, there came a point where some of these colonies were able to grow on chloramphenicol plates but unable to do so on ampicillin, possibly indicating a full curation of the plasmid. In order to confirm this, a new set of PCRs were carried out to confirm the presence of the Cm resistance gene.

34 colony PCRs for the whole Q $\beta$  fragment with the chloramphenicol resistance gene (4715 bp) were run using colonies from different experiments and while in some of them a band does appear, in several of them no band appears, indicating that the plasmid is no longer present in the cell, but despite that, the resistance somehow remains, further supporting the possibility that it does so in the form of RNA, although we cannot rule out other possibilities, such as a spontaneous mutation.

The results from this assay and a comparative between the colonies can be seen in Figure 46. Some of these selected colonies were made into glycerol stocks and kept at -80 °C for the laboratory's stocks, under the names SAJ738 for the original transformed strain and SAJ739 for the curated one.

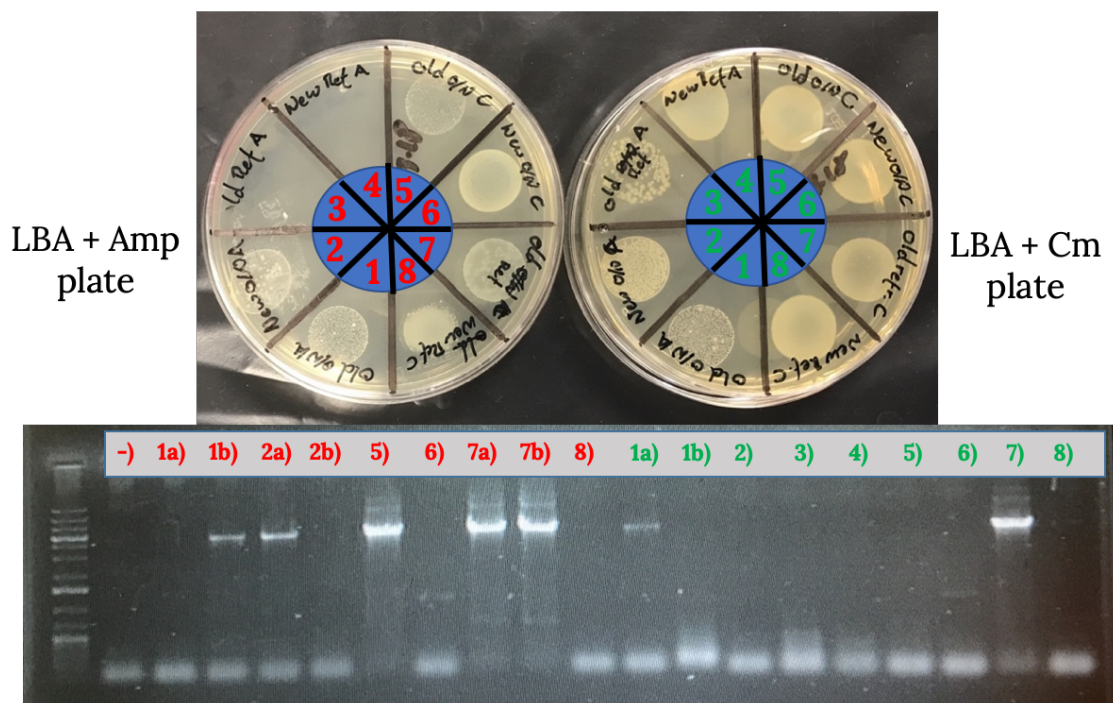


Figure 46. Colony growth and diagnostic PCRs for Q $\beta$  presence.



Comparison of eight different BL21 AI colonies grown in both Amp (red) and Cm (green) LB + agar plates. As seen in A), the growth in Amp is much sparser than in Cm plates, corresponding to a possible loss of Amp resistance over subsequent passages. B) Shows the result from PCR amplification of the same fragment of 4715 bp, corresponding to the set of genes from Q $\beta$ , in each of the different colonies. While some of them (mostly in the Amp plate) show a band at the expected size, the PCRs corresponding to the Cm-grown plates do not produce one. In some case PCR samples are labelled “a” or “b” since two different colony morphologies were present, and were tested independently.

In order to confirm the transitory nature of the cells' newly acquired chloramphenicol resistance due to the Q $\beta$  phage, a similar experiment was devised, so as to eliminate it over the course of several generations.

Cells would once again be cultured in liquid broth under two conditions; the presence of chloramphenicol and the absence of any antibiotics, and then cultured in 5 ml of LB in the absence of any antibiotics. The cells grown in the latter case would then be plated into regular LB agar plates and cultured once again in 5 ml of LB + Cm and just LB. Despite the cells being able to grow in both conditions for the first plating rounds, showing similar types of growth in both cases, by round 9 the bacteria seemed to have fully lost their ability to grow in presence of chloramphenicol. This was assessed both via plating and liquid culture, with the difference being easily discernible in both cases.

Glycerol stocks of all of the different steps were obtained and then used to estimate the number of colony-forming units (CFU) for each plating step under the two different conditions (presence and absence of chloramphenicol) at the same OD values. Multiple repetitions of this showed a much higher number of CFU in the absence of chloramphenicol, which in itself could be understandable, since the resistance is an exogenous gene and there is no antibiotic stress for the cells; but the ratio of CFU in the absence of chloramphenicol/CFU in presence of chloramphenicol greatly increases as the plating steps go by, coming to a point where there are essentially no colonies forming in the presence of the antibiotic.

These results, indicating the number of average CFUs per generation and their growth on LB agar plates can be seen in Figure 47. It indicates how the number of CFUs remained similar between the LB agar and LB agar + Cm media until a certain point where the latter starts progressively decreasing.

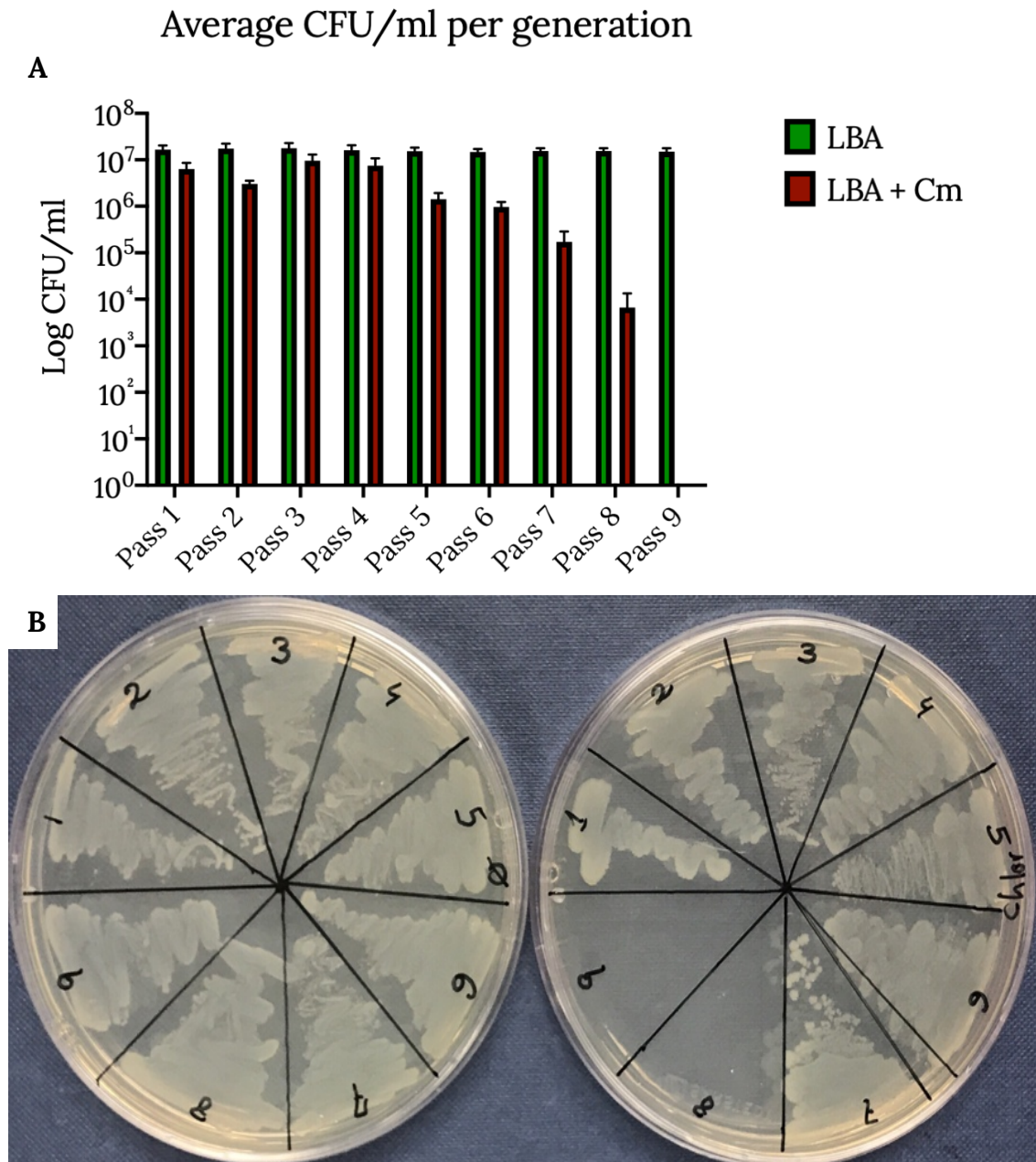


Figure 47. Comparative growth of generations of Q $\beta$ -expressing cells.

A) Average CFU values of the induced strains throughout the “curing” process of the Chloramphenicol Acetyltransferase gene carried within the Q $\beta$  phage’s genome; when plated in the absence of antibiotics or in the presence of chloramphenicol. Error bars represent the SEM. B) Streaks from each of the different colonies in LB agar plates (left) and LB agar + chloramphenicol (right). Error bars represent the Standard Error of the Mean (SEM).

#### 6.4.2 GFP experiments

In parallel, our collaborators at Sorbonne University in Paris assayed GFP expression via *in vitro* assays using the Protein synthesis Using Recombinant elements system, or PURE (Shimizu et al., 2001). This is a cell-free translation system containing a high-purified mixture of all necessary factors for protein synthesis, and which allows for a high purification rate of the obtained product. In this case, the assay took place using a new construct based upon our original Q $\beta$  plasmid, in which the *cat* gene was substituted for the *gfp* gene. This construct is described in Supplementary materials' Supplementary figure 10. Via qPCR, the preliminary measurements of GFP production were taken, and can be seen in Figure 48 comparing between a control sample with no GFP production and two different concentrations of the Q $\beta$ -GFP plasmid. The figures represent very early, preliminary data, and show that the plasmid constructs do indeed produce GFP, but at a slow rate, which stops at a certain point, indicating that the system needs to be improved.

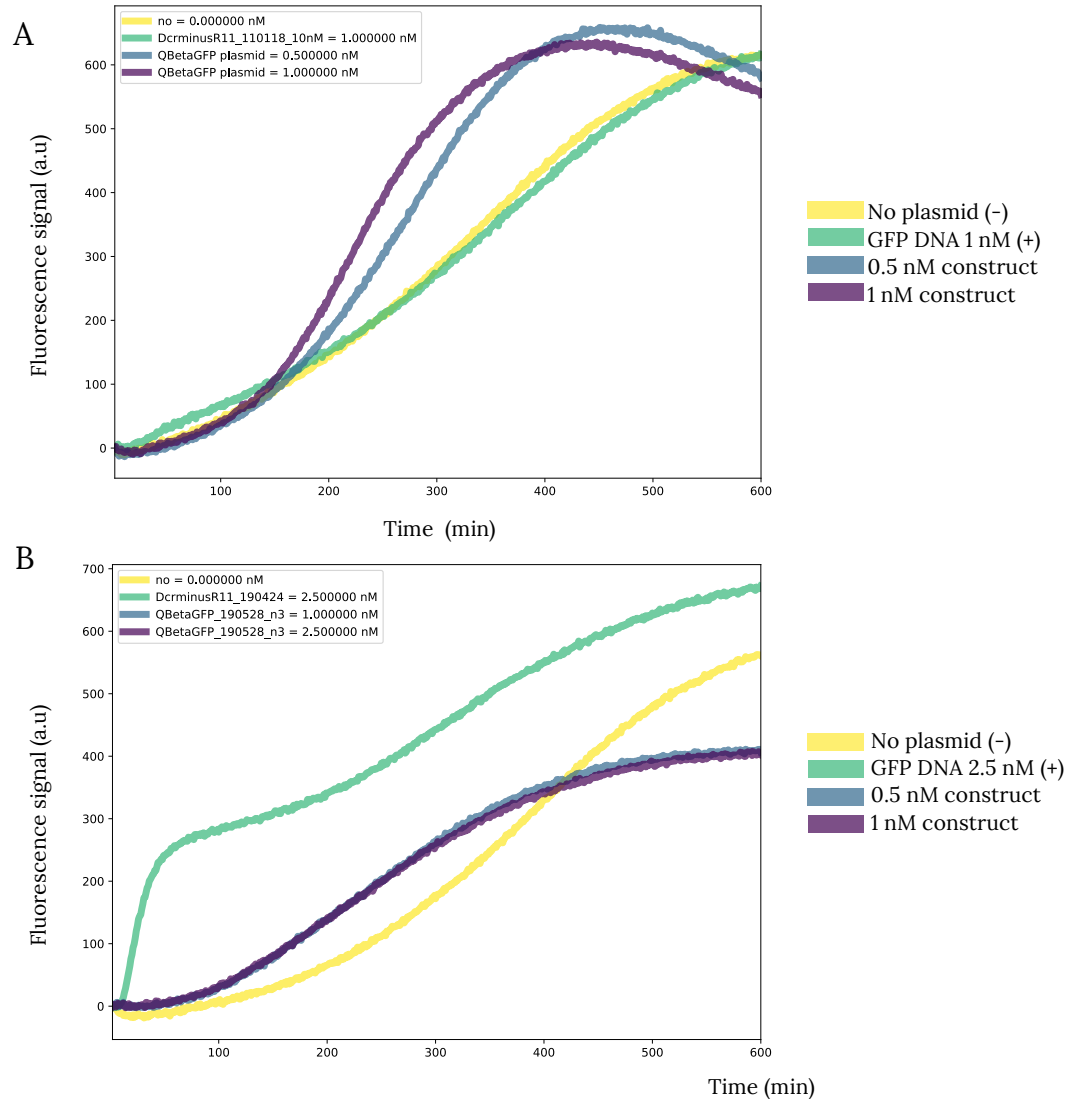


Figure 48. Q $\beta$ -GFP plasmid expression.

A) Fluorescence curves from qPCR representing preliminary results for GFP production in a PURE system by measuring fluorescence signal in AU (Y) over time in minutes (X). The different curves represent the PURE system without the plasmid (yellow), with DNA encoding for GFP as a control (green, which had issues rendering it unusable), and with different concentrations of the pUCIDT+Q $\beta$ +GFP plasmid (blue and violet). B) Second repetition of the assay, with a correct control DNA encoding for GFP (green), the PURE system without the plasmid (yellow), and once again, the two samples with different concentrations of pUCIDT+Q $\beta$ +GFP plasmid (blue and violet), which in this case show lower activation than in the previous assay.

## 6.5 Discussion of results

### 6.5.1 Q $\beta$ -cat experiments

The initial set of transformations made for these experiments worked in all cases, as each of the different strains was able to grow properly in the presence of ampicillin, and PCR confirmed the presence of the plasmid. However, the induction showed that each of the strains had a very different response to chloramphenicol. The different results between strains meant the elimination of two of them, as they were not responding properly to the induction (Figure 45). Both strains (*E. Coli* BL21 DE3 and LEMO DE3) had something in common, and that is their DE3 designation, which meant they contained the  $\lambda$ DE3 lysogen, carrying the gene for T7 RNA polymerase under control of the lacUV5 promoter, requiring IPTG to maximally induce expression. The only difference between them is the presence of the T7 lysozyme in LEMO DE3, designed to downregulate the expression of T7 RNA pol if needed, in cases where a specific protein was toxic for the cell.

This makes the results even more vexing, because LEMO DE3 shows more expression of the chloramphenicol resistance gene than BL21 DE3, despite having one extra level of regulation, the T7 lysozyme. If in both cases the induction system is the same, it is unclear why in one of the cases it does not seem to turn ON, not activating the antibiotic resistance, while in the other one it does the opposite, with the same resistance not being turned OFF, even when the T7 RNA polymerase is supposedly repressed by the lysozyme. Given this situation, and since the remaining strain (BL21 AI) seemed to work as desired when induced, it was then decided to use it for the remainder of the experiment.

This led to the results shown in Figure 46, where after several culture passages, the resistance to chloramphenicol is maintained, even when the original plasmid containing the gene encoding Cm resistance is not detectable via PCR. Our theory is that, given how the Q $\beta$  phage is a (+) ssRNA phage, the corresponding genetic information encoding chloramphenicol acetyltransferase is maintained in the form of RNA in the cytoplasm of the cell, not in plasmid form, as it was

previously. Coupled with the Q $\beta$ 's own replicase, which helps form an RNA-dependent RNA pol capable of replicating both (+) and (-) RNA strands, this creates a way for the information to remain unaffected by the bacterial defence system that is CRISPR, and ensure its presence in the cell, carried as part of the Q $\beta$  phage's genome.

In order to confirm that this resistance was only temporary and would only be present as long as it was needed, the same procedure used to eliminate the original plasmid was taken once again, (Figure 47). In here, the results indicate a slight difference when the cells are grown in the presence of the antibiotic, instead of in the absence of it, which could be explained by a possible deviation of resources to resistance production, but from the 5<sup>th</sup> pass onwards, we see a decline in number of CFUs, until a total disappearance in the final steps. This could be associated with either a decrease in the number of copies of the sequence carrying the resistance information, or in the production of the protein itself.

As further proof of the progressive disappearance of the chloramphenicol resistance, a retro-transcriptase qPCR assay was designed to properly assess the quantity of RNA present in each of the plating steps, in an attempt to associate those levels with the CFU results. Sadly, due to time constraints, such an assay did not take place in the end. However, it would be very relevant to do such an assay in order to elucidate the reason, and based on its results, decide on which assays would be more relevant afterwards.

#### 6.5.2 Qbeta-GFP experiments

The most important issue with this set of data is the fact that it is at a very early stage, and thus has several flaws and requires improvement. Our collaborators at Sorbonne University tested the GFP production of our Q $\beta$ -GFP plasmid at different concentrations in a PURE system against a control with no plasmid present, in order to assess the level of expression from the plasmids in the system, and provided us with the diagrams represented in Figure 48. The green curve acting as a positive control, failed to work in the first case due to an error

in the sample, but showed a very fast initial activation which then slowed down, but kept increasing over time. Meanwhile, the samples have different activation levels in the two experiments, but are both slow, weak, and seem to stop after 5 hours. One way to tackle this has been the increase of DNA concentration in the samples, to levels more comparable to the positive control. Sadly, no results are available as of yet. The gene fragment designed as a parallel experiment should also be present in these graphs both by itself and with the addition of the dsDNA pT7-RBS-rep. Doing this would show the differences in GFP expression once the replicase is produced and binds to the rev3'UTR\_MDV1. This should, in theory, produce GFP levels at a slower rate than the Q $\beta$ -GFP plasmids, and would act as a lower threshold control of sorts for the expression of the plasmid system.

This could also open the door to the question, why not use this system with a replicase *in trans* in an *in vivo* system instead of an *in vitro* one? Such a development would be interesting from a mechanistic point of view, and would certainly make things easier when designing systems, but it would be contradictory to the idea that spurred this project in the first place. Having a plasmid based around RNA and have it work only in bacterial strains that contain a replicase/RNA-dependent RNA polymerase would severely undercut the possibilities this system holds and its uses. Thus, we believe that, despite it making the system more complex, the current setup is a sufficiently effective way for inducible and transient gene expression that could exist outside the realm of current CRISPR systems, making a significant impact in the field of synthetic biology at large.

In conclusion, this data seems to confirm the following:

1. We have developed a working inducible system for transient gene expression based on the RNA phage Q $\beta$ , which regulated and maintained antibiotic resistance several generations after induction occurred.
2. More experiments are needed to improve and fully assess the workings of the system, including further repetitions of the GFP experiments, and RT-qPCR assays to properly assess the quantities of RNA in the system over generations after induction occurs.



## CHAPTER 7. OVERALL CONCLUSION & FUTURE DEVELOPMENTS

The main focus throughout the length of this thesis was the development of a procedure for the directed evolution of regulatory molecules via the use of T7 phages. While this procedure was focusing mainly in evolution of RNA elements, more specifically riboswitches, other alternative molecules have been presented in which such a strategy could be of use, such as proteins. The creation and testing of these different strategies have been recorded, along with the discussion of results dealing with their main strengths and possibilities, as well as current limitations.

Regarding the Main objective of the thesis, which revolved around setting up a system for the directed evolution of a riboswitch using T7 phage, many aspects need to be addressed to determine whether or not this has been achieved.

The starting point, as explained previously, was the development of a riboswitch library and its subsequent incorporation into plasmids and phages for their use; which was dealt with in Chapter 3. Next, it was the assembly and testing of strains used for the positive and negative selection of these riboswitches, and their use to assemble an evolutionary system using all these elements, which was discussed in Chapter 4. Chapter 4 also deals with the results from the direct evolution of riboswitches using the newly created system.

### 7.1 Development of a phage riboswitch library

The results obtained in Chapter 3 confirm that the library-assembling strategy used did indeed allow for the creation of a library of 8N randomised riboswitches, which was introduced into a T7 phage population via homologous recombination. Theoretically, a library of eight-nucleotide fragments means  $4^8$ , that is 65536 different variants.

This was then fully confirmed via the Next Generation Sequencing results present in Chapter 4, which at the same time confirmed an absence of bias in the library,



since the obtained results show no particular sequence has prominence when compared to others.

This gave us enough certainty to assume all different riboswitch variants containing the random 8N sequence were present in the initial library before the evolutionary process started.

### 7.2 Development of strains for positive and negative selection of riboswitches

The positive and negative selection strains, whose development was described within Chapter 4, were tested appropriately, via plaque, spot, one-step and growth assays, in order to identify the best candidate for each selection step. However, both candidates for the positive and negative selections required improvements in order for their selection threshold to be optimal, insofar as it would only be regulated by the activation or inactivation of the riboswitch.

The first element that should be addressed regarding the flaws of the selection process is the promoter leakage. The structure of the pRM promoter itself allows for multiple different ways where leakage happens, (Vilar & Saiz, 2010), and this certainly impacts the efficiency of selection. There is a difference between the presence or absence of theophylline in the system in both the positive and negative selections, but the degree in which that difference is caused solely by the activation of the riboswitch is not properly known, as the leakage of the selection genes influence the results. Leakage levels need to be decreased in order for selection to be more stringent and efficient.

Although leakage affects both selection strains, the negative selection step shows a much smaller difference between the ON and OFF states of the riboswitch. The cause of it is unclear and is also what drives home another big limitation of this study: the number of unknowns in the process by which the T7-exclusion protein pifA blocks T7 phage replication. In order to improve the negative selection step, more information needs to be obtained about the molecular pathways pifA is involved in. Since this falls outside the scope of the project, another possibility

could be the measurement of pifA protein levels both when the riboswitch is OFF and when it is ON. This could be done using different protein measuring techniques, such as a western blot to detect pifA levels specifically, or a Bradford assay to detect overall cell and T7 protein concentrations in the sample and assess any differences. The first technique would allow us to more accurately measure the difference in pifA production between both states, and possibly correlate that to riboswitch activation.

### 7.3 Development of an evolutionary system

The development of the evolution system using all these elements changed throughout the development of the project, mainly due to the technical difficulties arising after the breaking of the laboratory's in-house-developed bioreactor, which forced the need to produce an alternative procedure that was not as automatic as the original one. If fixed, it would be interesting to compare the results obtained via the originally planned procedure with the ones obtained throughout this PhD and consider any differences and their relation to the method used in that particular case.

### 7.4 Direct evolution of a riboswitch

Regarding the considered main interest of the project, that is, the directed evolution of the riboswitch, the data presented in Chapter 4 can argue in favour of that, since the NGS data indicates a drastic reduction of sequence variation over subsequent generations, going from a state where no sequence has prominence over the rest to one where one sequence is substantially overrepresented in the population. Virulence factor experiments could be used to support this conclusion, as we can see that the values change over generations and theophylline concentrations; but again, the separation between a phage population's virulence and riboswitch activation needs to be taken into account. Further assessments into the production levels for each of the proteins managing the selection could help clear out this issue.

Despite its shortcomings, this project has managed to reach the objectives it set out to fulfil, and although it can be improved in several aspects, this should not overshadow the impact that such a strategy could have in the field of riboswitch development and improvement, and tentatively, biosensor research at large. As far as we know, this research represents the first instance of a riboswitch evolution strategy via the use of T7 phage, and the first time a working riboswitch has been included in the genome of a phage.

### 7.5 Secondary objectives:

As for the secondary objectives, due to time constraints, several unanswered questions remain, but the way to follow through is clear:

Regarding the evolution of alternative sensors, the system needs to be improved, and in order to do so, once again, the issues with pifA need to be addressed, probably with the same protein production measurements recommended for the main objective. The evolutionary procedure should be fully tested in the same manner as the riboswitch strategy, with the presence or absence of the molecule of interest, confirming over generations whether or not the system manages to sway the protein of interest towards the desired objective, an improved protein receptor with higher orthogonality.

As for the alternative strategy for evolution based on the production of phagemid particles, the main issues correspond to the fact that the evolutionary procedure needs to be setup properly in order to confirm whether or not it works over generations.

If it does indeed work as expected, it would be certainly interesting to compare these results with the ones obtained from the full phage strategy and assess if the same sequences were selected. If this were not the case, comparing the resulting sequences and understanding how they came to be selected could deepen our knowledge about the system by which riboswitches are selected, based on their structural-functional characteristics.

Finally, the results obtained from the development of the Q $\beta$  phage-based inducible system seem to be ones that work according to what was expected, showing a working inducible activation that remains over several generations but can be eliminated when it is not necessary anymore. The possibilities that such a system offers could make a big impact in the field of microbiological research, with possible uses including more tightly regulated industrial processes, bioremediation applications, or even possible uses in gene therapy.

### 7.6 Closing Remarks:

As a closing statement, over the course of this PhD, the main research goal was concerned with the development of a system starting from a known riboswitch with a randomised fragment, but it could even lead to the evolution of new switches from random sequences. New switches that could bind known molecules and teach us more about the structural basis for their function, or even switches that react to newly developed compounds, with all the possibilities that offers.

This system itself evolved, serving as a starting point for different alternatives that, while working with different structures, ultimately have the same objective, the development and improvement of biosensor molecules in an *in vivo* environment.

Strategies like this could mean improvements not just at a basic research level, in the form of future riboswitch structure and development research, but in the applied research level as well. The development of improved or novel receptors could spearhead the development of new, more efficient processes such as the production of industry-relevant compounds with increased efficiency and yield, personalised medical treatments and a new way to fight the current antibiotic crisis. Even the possibility of developing new biosensors and microorganisms for bioremediation strategies, capable of degrading contaminants in the environment in an efficient way and shut down after fulfilling their objective.

As this thesis is being written, the proverbial bricks for some of these projects are just being laid, with years standing between us and the point where they will become a reality and bear fruit, being able to help society. But that does not mean work like this is any less important because of it.

While applied research is important and gives faster benefits; basic research, despite its waiting, meandering, its non-profitability, and its difficulty, lays down the foundations for it. It opens up the road to further scientific innovation, and towards new possibilities we can't yet see.

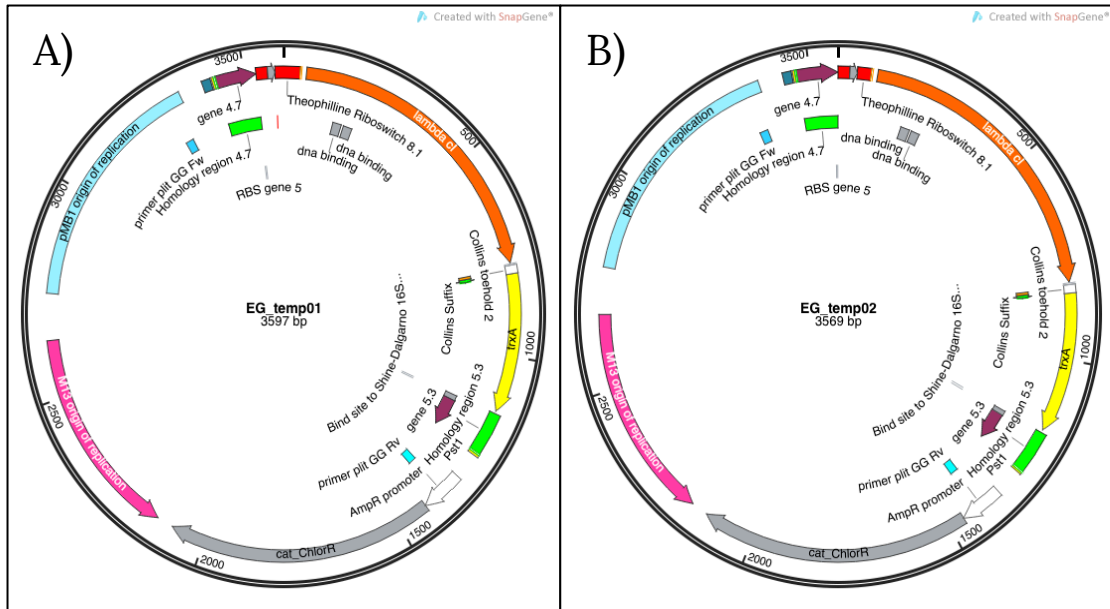
Including, hopefully, a better future.

千里之行, 始於足下 : A journey of a thousand miles starts beneath one's feet

“It's a dangerous business, Frodo, going out your door. You step onto the road, and if you don't keep your feet, there's no knowing where you might be swept off to.” (Tolkien, 1954)

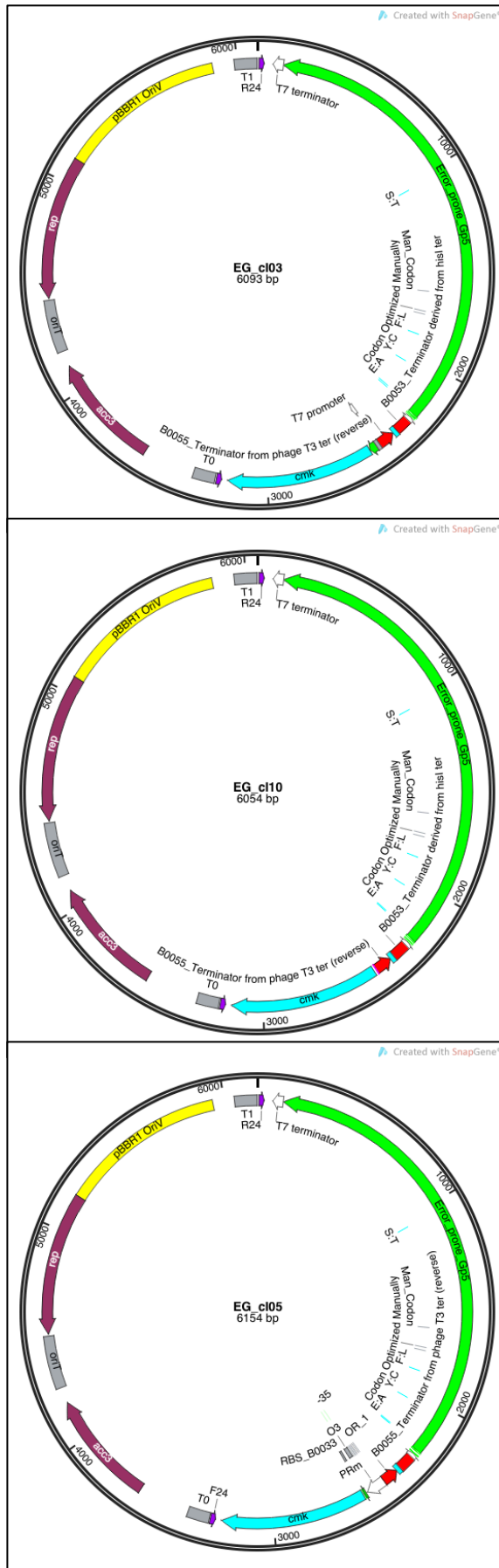
## 8 SUPPLEMENTAL MATERIALS:

All plasmids described throughout this thesis are described in this section and can be found in the Laboratory's LIMS database and sample collection.



Supplementary figure 1. Intermediate recombinant plasmid constructs.

Temporary sequences produced from the modification of EG\_001 in order to obtain a plasmid library of riboswitch variants. A) Map of the plasmid resulting from the amplification of EG\_001 with primers EG\_GGFw and EG\_GGRev, incorporating several genetic elements between gene 4.7 and the  $\lambda$  cI gene. B) Map of the resulting plasmid produced after BpiI/BbsI digestion and Gibson assembly, which will be used for the incorporation of the fixed version of the riboswitch and the library of different riboswitch sequences.

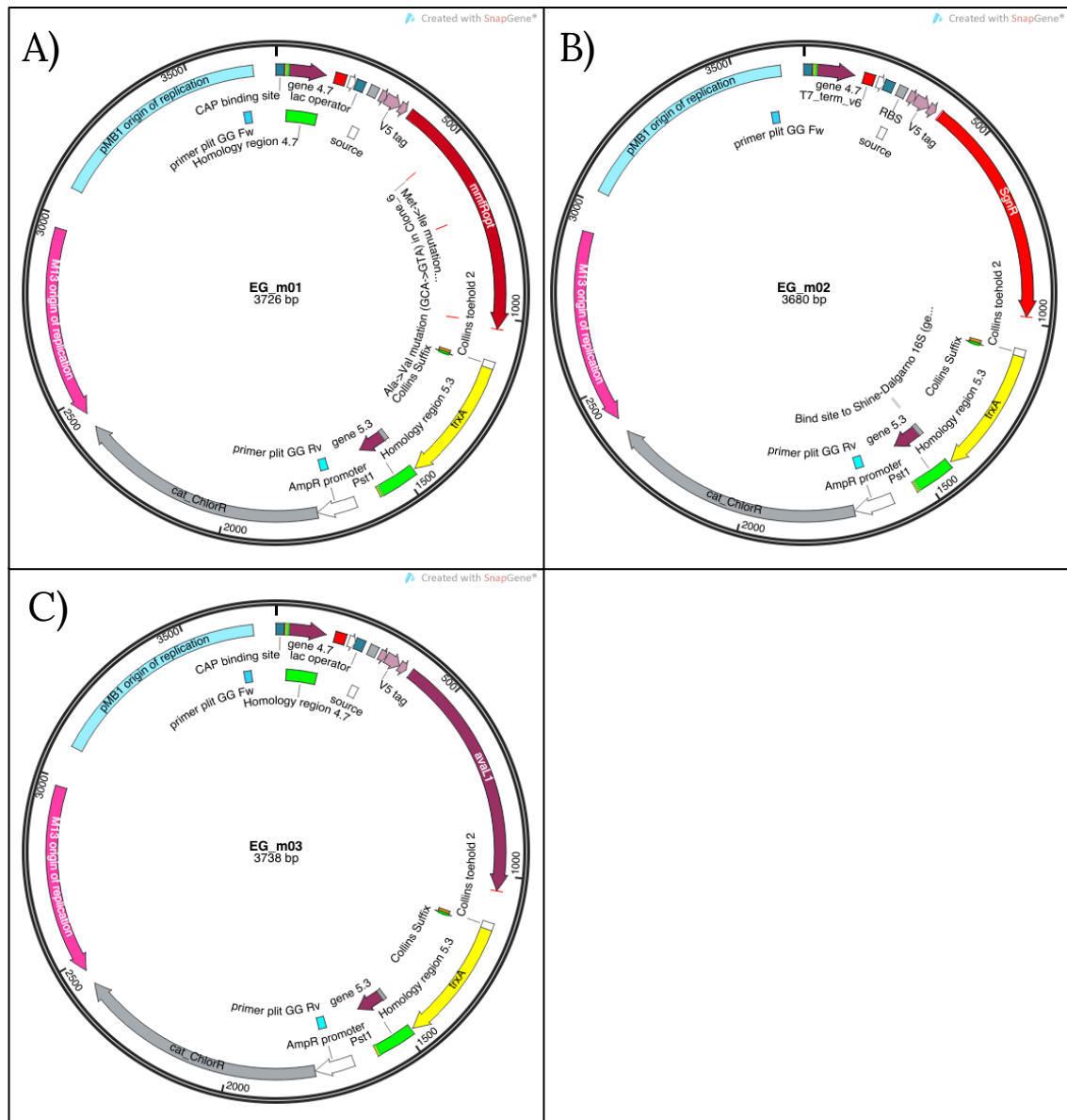


Supplementary figure 2. Positive control plasmid for positive selection.

Map of plasmid EG\_cI03, with the most relevant elements for the positive selection indicated by name. T7 promoter regulating expression of cmk gene, ensuring high expression during T7 infection.

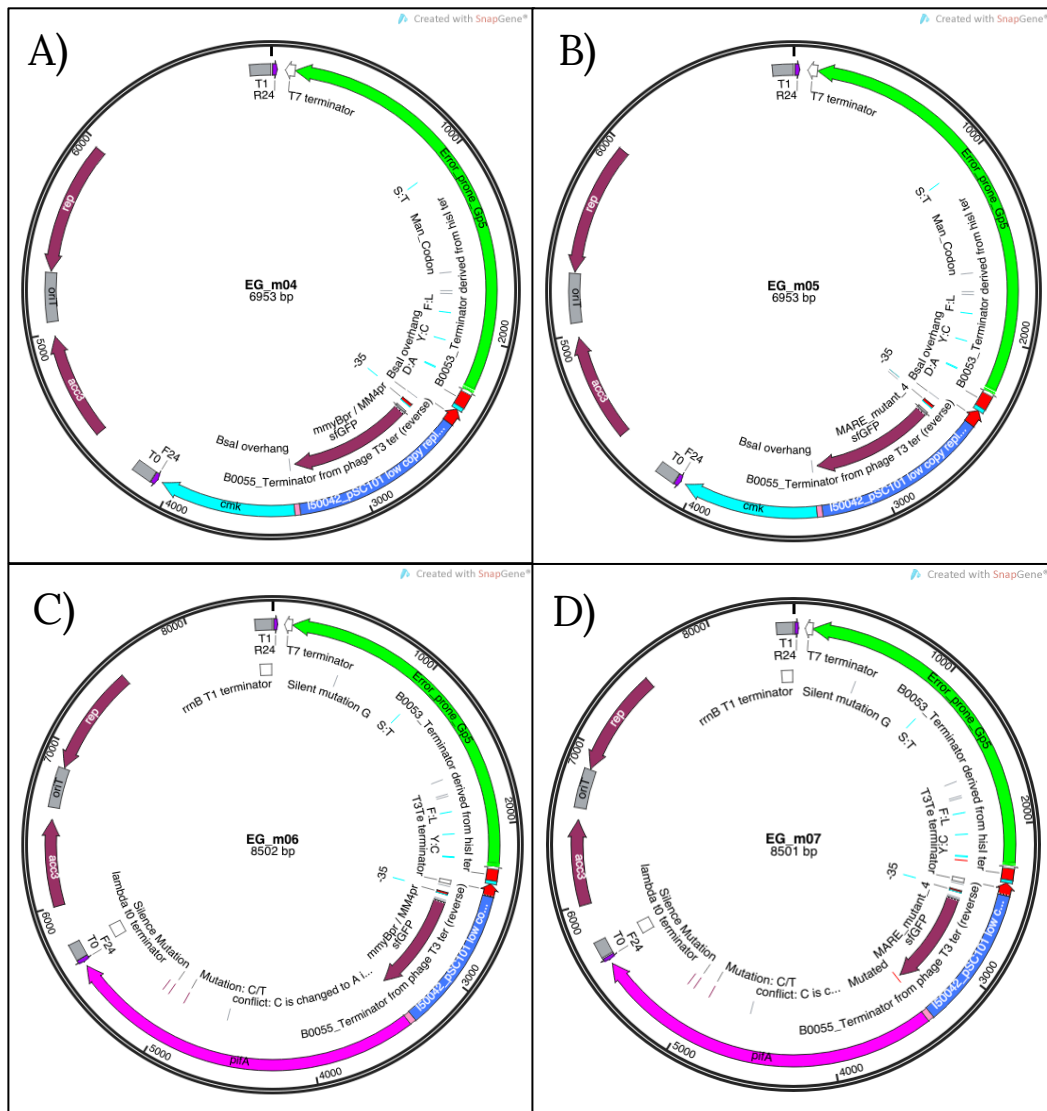
Maps corresponding to the plasmids used for the negative selection testing, problem plasmids as well as controls. A) Positive control plasmid EG\_c11, with a T7 promoter regulating *pifA* expression. B) Negative control plasmid EG\_c118, with no promoter nor RBS regulating *pifA* expression. C) Test plasmid EG\_c115, used as an example for all 6 test plasmids of the same size and structure.





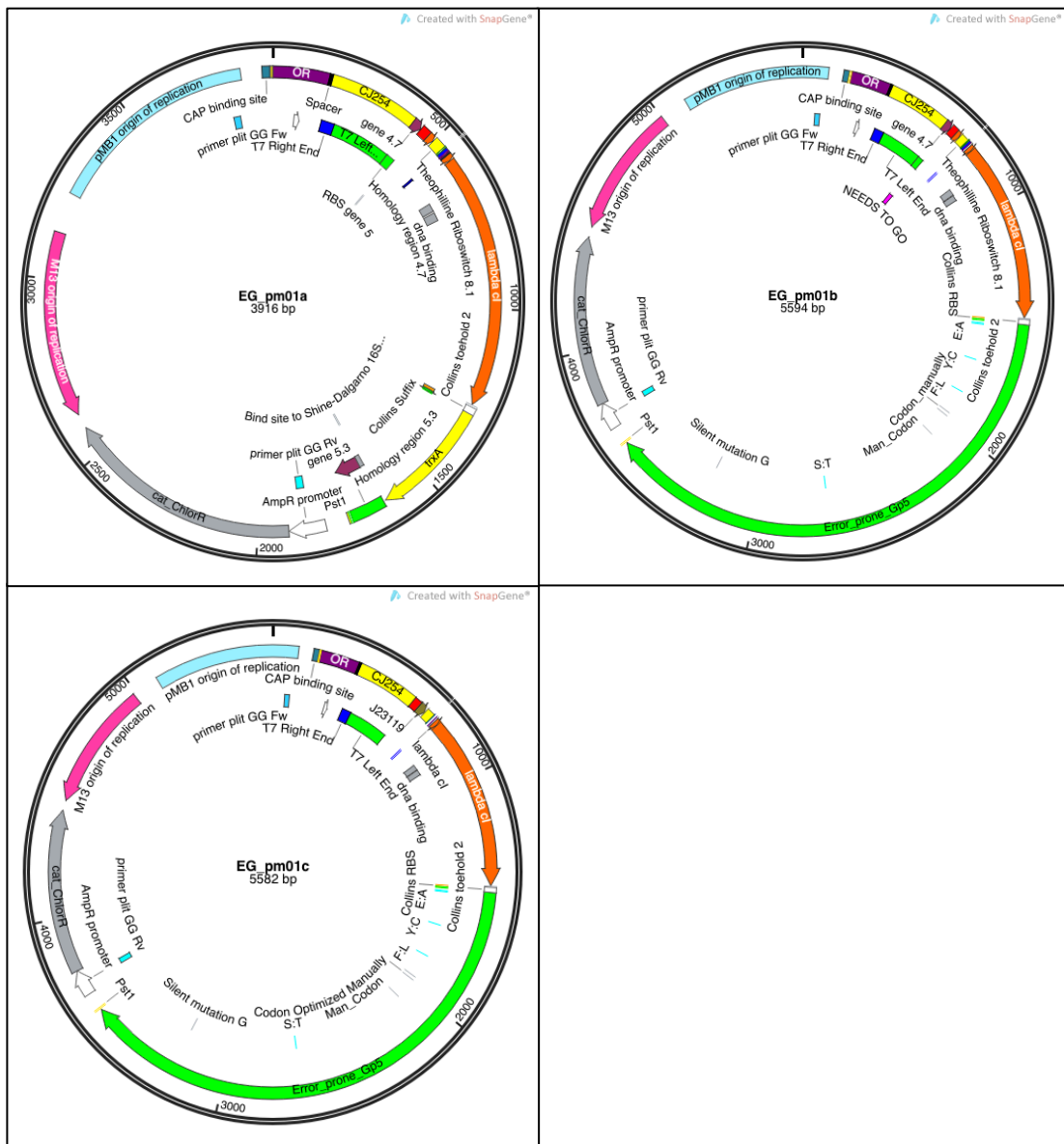
Supplementary figure 6. Homologous recombination plasmids for *mmfR* variants.

Plasmids map for the homology recombination procedure in the mmfR strategy showcasing the different elements present, especially the ones within the homology region, which are to be introduced within the phage for the evolution process. A) Map of plasmid EG\_m01, containing the original mmfR sequence. B) Map of plasmid EG\_m02, containing the alternative sgnR sequence. C) Map of plasmid EG\_m03, containing the alternative avaL1 sequence.



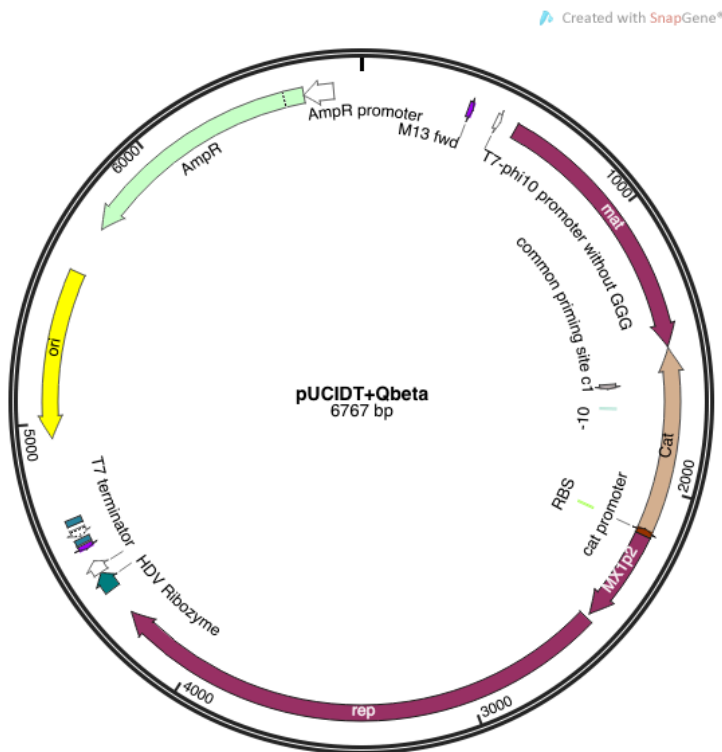
Supplementary figure 7. Selection plasmids for the *mmfR* strategy.

Backbones designed for the projects' selection procedures. A) Negative selection plasmid A, carrying the WT *mmyB* operator regulating sfGFP and *cmk* expression. B) Negative selection plasmid B, carrying the mutated MARE4 version of the operator regulating the same genes. C) Positive selection plasmid A, carrying sfGFP and *pifA* under the control of the WT *mmyB* operon. D) Positive selection plasmid B, carrying the same genes, with the difference of the mutated MARE4 operator.



Supplementary figure 8. Phagemid plasmid versions.

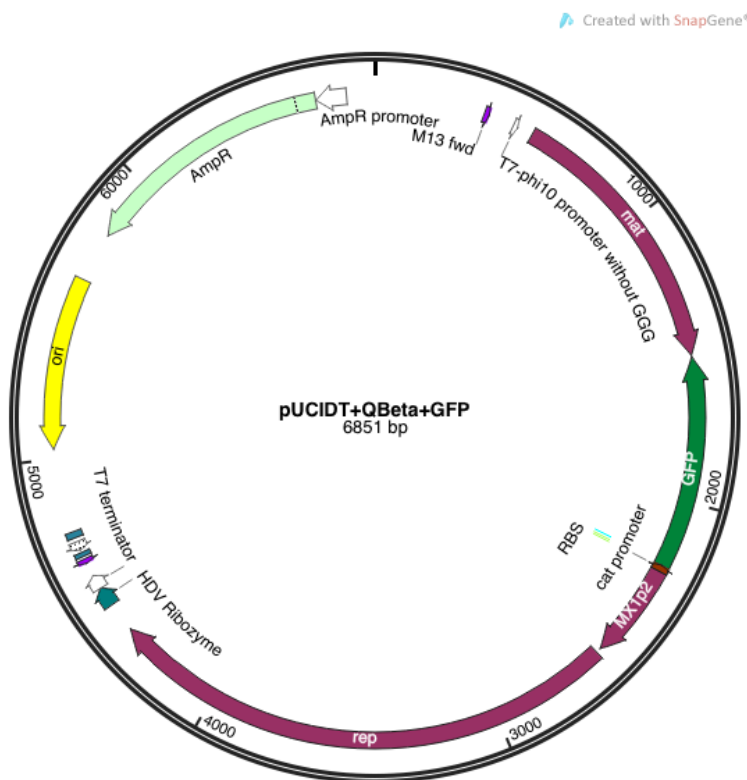
Plasmid maps of the three different constructs gradually assembled as part of the strategy. As shown, pm01a is very similar to the plasmid used for the homologous recombination strategy, but gene 4.7 has been substituted by the synthetic T7 packaging signal sequence. Pmb01b includes the addition of the error-prone gp5, and pmb01c finally eliminates the T7 promoter in front of the riboswitch in favor of an *E. coli* one, as well as eliminates the remnants of the homology region with gene 4.7.



Supplementary figure 9.

Original Q $\beta$  IDT plasmid.

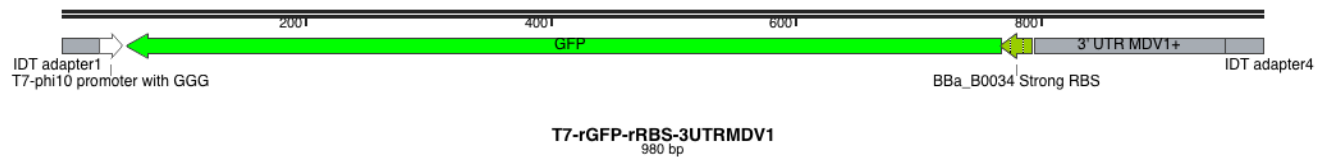
Map of the IDT-assembled construct pUC\_cat, formed of 67667 bp, with a pUC plasmid backbone from IDT and the Q $\beta$  phage's genes in the 5'-3' orientation, as well as the chloramphenicol acetyltransferase gene in a 3'-5' orientation.



Supplementary figure 10.

Q $\beta$ -GFP plasmid.

Map of the lab-assembled construct pUC-gfp, based upon and with the same elements as pUC-cat, but with a gfp gene in a 3'-5' orientation, making it 6851 bp in total.



Supplementary figure 11 . 3'UTR gene fragment for PURE experiments.

Structure of the gene fragment assembled by IDT, containing an adapter, a T7 promoter in a 5'-3' sense, and a 3'UTR, RBS and GFP sequence in 3'5' sense.

Class	Function <sup>a</sup>
<b>Class I</b>	
0.3	B-DNA mimic; anti-type I restriction
0.4, 0.5, 0.6A, 0.6B	Not conserved; nonessential
0.7	Protein kinase; host-transcription shutoff; Col Ib exclusion 84, 211
211	
1	T7 RNA polymerase
1.1	Conserved, nonessential
1.2	<i>E. coli</i> dGTPase inhibitor; F-exclusion
1.3	DNA ligase
<b>Class II</b>	
1.4, 1.5, 1.6	Not conserved, nonessential
1.7	Full-length gene not conserved; beneficial for growth
1.8	Poorly conserved, nonessential
2	<i>E. coli</i> RNAP inhibitor
2.5	SSB
2.8	Not conserved, nonessential; homing endonuclease?
3	Endonuclease I, Holliday junction resolvase
3.5	Amidase (lysozyme); regulates T7 RNAP activity
3.8	Not conserved, nonessential; homing endonuclease
4A	Primase-helicase; gp4B helicase from internal in-frame start
4.1, 4.2	Overlapps; not conserved
4.3, 4.5	Conserved, nonessential
4.7	Not conserved, nonessential
5	DNA polymerase
5.3	Not conserved, nonessential, homing endonuclease
5.5	Conserved, nonessential, binds <i>E. coli</i> HNS; $\lambda$ rex exclusion
	Non-conserved -1 frameshift leads to T7 5.5-5.7 fusion
5.7	Conserved, nonessential
5.9	Inhibits RecBCD nuclease, nonessential, not conserved
6	5'→3' double-stranded exonuclease, RNase H
6.3	Poorly conserved, nonessential
<b>Class III</b>	
6.5	Conserved, nonessential
6.7	Virion protein; ejected into infected cell
7	Nonessential, not conserved; host range
7.3	Essential virion protein; ejected into infected cell
7.7	Not conserved; homing endonuclease
8	Head-tail connector protein
9	Scaffolding protein
10A	Major capsid protein; -1 frameshift yields minor capsid protein gp10B F exclusion
11	Tail protein
12	Tail protein
13	Essential; required for gp6.7 incorporation in virion
14	Internal core protein; ejected into infected cell
15	Internal core protein; ejected into infected cell
16	Internal core protein; ejected into infected cell
17	Tail fiber protein
17.5	Class II holin
18	Small terminase subunit
18.5-18.7	Conserved; $\lambda$ Rz-Rz1 homologs
19	Large terminase subunit
19.2, 19.3	Overlapps, conserved
19.5	Nonessential, conserved

<sup>a</sup>Conserved or not conserved refers to close relatives of T7.

Supplementary Table 1 . List of T7 genes

From (Molineux, 2005). Full list of T7 genes and their functions

Supplementary Table 2. Primer combinations for each NGS sample.

Matrix indicating which primers correspond to each specific phage library in each of the two repetitions. The 6x4 matrix gives the possibility of separating between 24 samples at once, each of them with their own specific signal, thanks to the specific indexes present in each of the 2 primers used for the amplification. The different denominations for each library are based on the selection step they were obtained from, followed by the number of the generation. If there are multiple libraries corresponding to the same generation, a letter or date is included after the denomination. E.g. Sample cmk11 corresponds to the phage library obtained after the 11<sup>th</sup> positive selection step in the evolutionary process. A and B correspond to the 1<sup>st</sup> and 2<sup>nd</sup> set of sequenced libraries, respectively. B is not currently represented in the results due to a lack of time but will be present in the final version of this thesis.

A)	N701	N702	N703	N704	N705	N706
S502	PLASMID Pool	Fix Riboswitch	Rd Switch 1	Rd Switch 2	Cmk1	cmk2
S503	pifA2	cmk5	pifA6	cmk7	pifA7	pifA9
S504	pifA10	cmk11	pifA11	cmk12	pifA12	pifA12B
S517	pifA13	pifA13D	pifA13B_8-11	cmk14	pifA14_6-11	pifA14
B)	N701	N702	N703	N704	N705	N706
S502	Rd1	Rd2	Rd3	Cmk1	pifA1	cmk3
S503	pifA3	cmk4	pifA4	cmk5	pifA5	cmk6
S504	cmk8	pifA8	cmk9	cmk10	pifA11	cmk12
S517	cmk13	cmk15	pifA15	cmk 16	cmk2 Ampl	cmk16 Ampl

Supplementary Table 3. qPCR plate distribution of samples.

Distribution of library and standard samples in the qPCR plate. 13 out of all the 24 PCR-amplified samples were taken in order to check the presence of to specific Illumina tags required for MiSeq detection. Each sample was represented in the form of two dilutions, except for the first one, which had three to make the most use of the plate. Each dilution consisted of three repetitions in order to ensure statistical significance of the results.

	1	2	3	4
A	Standard 1 10pM	Standard 1 10pM	Standard 1 10pM	Library 2 $10^{-4}$
B	Standard 2 1 pM	Standard 2 1 pM	Standard 2 1 pM	Library 2 $10^{-5}$
C	Standard 3 0.1 pM	Standard 3 0.1 pM	Standard 3 0.1 pM	Library 3 $10^{-4}$
D	Standard 4 0.01 pM	Standard 4 0.01 pM	Standard 4 0.01 pM	Library 3 $10^{-5}$
E	NTC	NTC	NTC	Library 4 $10^{-4}$
F	Library 1 $10^{-3}$	Library 1 $10^{-3}$	Library 1 $10^{-3}$	Library 4 $10^{-5}$
G	Library 1 $10^{-4}$	Library 1 $10^{-4}$	Library 1 $10^{-4}$	Library 5 $10^{-4}$
H	Library 1 $10^{-5}$	Library 1 $10^{-5}$	Library 1 $10^{-5}$	Library 5 $10^{-5}$
	5	6	7	8
A	Library 2 $10^{-4}$	Library 2 $10^{-4}$	Library 6 $10^{-4}$	Library 6 $10^{-4}$
B	Library 2 $10^{-5}$	Library 2 $10^{-5}$	Library 6 $10^{-5}$	Library 6 $10^{-5}$
C	Library 3 $10^{-4}$	Library 3 $10^{-4}$	Library 7 $10^{-4}$	Library 7 $10^{-4}$
D	Library 3 $10^{-5}$	Library 3 $10^{-5}$	Library 7 $10^{-5}$	Library 7 $10^{-5}$
E	Library 4 $10^{-4}$	Library 4 $10^{-4}$	Library 8 $10^{-4}$	Library 8 $10^{-4}$
F	Library 4 $10^{-5}$	Library 4 $10^{-5}$	Library 8 $10^{-5}$	Library 8 $10^{-5}$
G	Library 5 $10^{-4}$	Library 5 $10^{-4}$	Library 9 $10^{-4}$	Library 9 $10^{-4}$
H	Library 5 $10^{-5}$	Library 5 $10^{-5}$	Library 9 $10^{-5}$	Library 9 $10^{-5}$
	9	10	11	12
A	Library 6 $10^{-4}$	Library 10 $10^{-4}$	Library 10 $10^{-4}$	Library 10 $10^{-4}$
B	Library 6 $10^{-5}$	Library 10 $10^{-5}$	Library 10 $10^{-5}$	Library 10 $10^{-5}$
C	Library 7 $10^{-4}$	Library 11 $10^{-4}$	Library 11 $10^{-4}$	Library 11 $10^{-4}$
D	Library 7 $10^{-5}$	Library 11 $10^{-5}$	Library 11 $10^{-5}$	Library 11 $10^{-5}$
E	Library 8 $10^{-4}$	Library 12 $10^{-4}$	Library 12 $10^{-4}$	Library 12 $10^{-4}$
F	Library 8 $10^{-5}$	Library 12 $10^{-5}$	Library 12 $10^{-5}$	Library 12 $10^{-5}$
G	Library 9 $10^{-4}$	Library 13 $10^{-4}$	Library 13 $10^{-4}$	Library 13 $10^{-4}$
H	Library 9 $10^{-5}$	Library 13 $10^{-5}$	Library 13 $10^{-5}$	Library 13 $10^{-5}$



## 9 BIBLIOGRAPHY

- Abou-Nader, M., & Benedik, M. J. (2010). Rapid generation of random mutant libraries. *Bioengineered Bugs*, 1(5), 337–340.
- Ackerman, H. W. (2009). *Phage Classification and Characterization*. In: *Bacteriophages, Methods in Molecular Biology*. (M. R. Clokie & A. M. Kropinski, Eds.). Clifton, NJ: Humana Press.
- Adams, M. H. (1959). *Bacteriophages*. London: Interscience Publishers Ltd., London.
- Alam, T. I., Draper, B., Kondabagil, K., Rentas, F. J., Ghosh-Kumar, M., Sun, S., ... Rao, V. B. (2008). The headful packaging nuclease of bacteriophage T4. *Molecular Microbiology*, 69(5), 1180–1190.
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Anonymous. (2010). M9 Minimal Medium (standard). *Cold Spring Harbor Protocols*. <https://doi.org/10.1101/pdb.rec12295>.
- Anzalone, A. V, Lin, A. J., Zairis, S., Rabadan, R., & Cornish, V. W. (2016). Reprogramming eukaryotic translation with ligand-responsive synthetic RNA switches. *Nature Methods*, 13(5).
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., ... Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Molecular Systems Biology*, 2.
- Barnes, P. J. (2013). Theophylline. *American Journal of Respiratory and Critical Care Medicine*, 188(5), 901–906. <https://doi.org/10.1164/rccm.201302-0388PP>
- Barsacchi, M., Novoa, E. M., Kellis, M., & Bechini, A. (2016). Structural bioinformatics SwiSpot : modeling riboswitches by spotting out switching sequences. *Bioinformatics*, 32(21)(July), 3252–3259.
- Beisel, C. L., & Smolke, C. D. (2009). Design Principles for Riboswitch Function. *PLoS Computational Biology*, 5(4), 1–14.
- Bengert, P., & Dandekar, T. (2004). Riboswitch finder — a tool for identification of riboswitch RNAs. *Nucleic Acids Research*, 32, 154–159.



- Bhalla, N., Jolly, P., Formisano, N., & Estrela, P. (2016). Introduction to biosensors. *Essays in Biochemistry*, 60(1), 1–8.
- Blount, K. F., & Breaker, R. R. (2006). Riboswitches as antibacterial drug targets. *Nature Biotechnology*, 24(12), 1558–1564.
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), 852–857.
- Borujeni, A. E., Channarasappa, A. S., & Salis, H. M. (2013). Translation rate is controlled by coupled trade-offs between site accessibility , selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research*, (23), 1–14.
- Borujeni, A. E., Mishler, D. M., Wang, J., Huso, W., & Salis, M. (2015). NAR Breakthrough Article Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic*, 8(44(1)), 1–13.
- Bradley, R. W., & Wang, B. (2015). Designer cell signal processing circuits for biotechnology. *New Biotechnology*, 32(6), 635–643.
- Braga, R. M., Dourado, M. N., & Araújo, W. L. (2016). Microbial interactions: ecology in a molecular perspective. *Brazilian Journal of Microbiology*, 47, 86–98.
- Breaker, R. R., Banerji, A., & Joyce, G. F. (1994). Continuous in Vitro Evolution of Bacteriophage RNA Polymerase Promoters. *Biochemistry*, 33(39), 11980–11986.
- Briand, L., Marcion, G., Kriznik, A., Heydel, J. M., Artur, Y., Garrido, C., ... Neiers, F. (2016). A self-inducible heterologous protein expression system in *Escherichia coli*. *Scientific Reports*, 6(December 2015), 1–11.
- Brown, D., & Gold, L. (1996). RNA replication by Q $\beta$  replicase: A working model. *Proceedings of the National Academy of Sciences of the United States of America*, 93(21), 11558–11562.
- Brown, R., Lengeling, A., & Wang, B. (2017). Phage engineering: how advances in molecular biology and synthetic biology are being utilized to enhance the therapeutic potential of bacteriophages. *Quantitative Biology*, 5(1), 42–54.

- Bunting, J. (1997). *BBC Horizon: The Virus that Cures*. United Kingdom: BBC.
- Carlson, J. C., Badran, A. H., Guggiana-nilo, D. A., & Liu, D. R. (2014). Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chemical Biology*, 10(March), 216–222.
- Carpenter, A. C., Paulsen, I. T., & Williams, T. C. (2018). Blueprints for biosensors: Design, limitations, and applications. *Genes*, 9(8).
- Caspersson, T., & Schultz, J. (1939). Pentose Nucleotides in the Cytoplasm of Growing Tissues. *Nature*, 143, 668–669.
- Ceres, P., Trausch, J. J., & Batey, R. T. (2013). Engineering modular ‘ON’ RNA switches using biological components, *Nucleic Acids Research*, 41(22), 10449–10461.
- Chan, B. K., Turner, P. E., Kim, S., Mojibian, H. R., Elefteriades, J. A., & Narayan, D. (2018). Phage treatment of an aortic graft infected with *Pseudomonas aeruginosa*. *Evolution, Medicine and Public Health*, 2018(1), 60–66.
- Chan, L. Y., Kosuri, S., & Endy, D. (2005). Refactoring bacteriophage T7. *Molecular Systems Biology*, 1(1), E1–E10.
- Chen, T., & Romesberg, F. E. (2014). Directed Polymerase Evolution. *FEBS Letters*, 588(2), 219–229.
- Cheng, X., Wang, W., & Molineux, I. J. (2004). F exclusion of bacteriophage T7 occurs at the cell membrane. *Virology*, 326(2), 340–352.
- Chi, X., Zhang, S., Sun, H., Duan, Y., Qiao, C., Luan, G., & Lu, X. (2019). Adopting a Theophylline-Responsive Riboswitch for Flexible Regulation and Understanding of Glycogen Metabolism in *Synechococcus elongatus* PCC7942. *Frontiers in Microbiology*, 10, 551. Retrieved from <https://www.frontiersin.org/article/10.3389/fmicb.2019.00551>
- Chung, Y., & Hinkle, D. C. (1990a). Bacteriophage T7 D N A Packaging II . Analysis of the DNA Sequences Required for Packaging Using a Plasmid Transduction Assay. *Journal of Molecular Biology*, 216, 927–938.
- Chung, Y., & Hinkle, D. C. (1990b). Bacteriophage T7 DNA Packaging I . Plasmids Containing a T7 Replication Origin and the T7 Concatemer Junction are Packaged into Transducing Particles During Phage

- Infection. *Journal of Molecular Biology*, 216, 911–926.
- Cobb, R. E., Sun, N., & Zhao, H. (2013). Directed Evolution as a Powerful Synthetic Biology Tool. *Methods*, 60(4), 81–90.
  - Corre, C., Song, L., O'Rourke, S., Chater, K. F., & Challis, G. L. (2008). 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *Streptomyces coelicolor* genome mining. *Proceedings of the National Academy of Sciences*, 105(45), 17510–17515.
  - Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, 227(6), 561–563.
  - Cui, W., Cheng, J., Miao, S., Zhou, L., Liu, Z., Guo, J., & Zhou, Z. (2017). Comprehensive characterization of a theophylline riboswitch reveals two pivotal features of Shine-Dalgarno influencing activated translation property. *Applied Microbiology and Biotechnology*, 101(5), 2107–2120.
  - D'Herelles, F. (1949). The Bacteriophage. *Science News*, (14), 44–59.
  - Davson, H., & Perkins, E. S. (2018). Human Eye. In *Britannica*. Encyclopædia Britannica, inc.
  - Delbrück, M. (1946). Bacterial Virus or bacteriophages. *Biological Review of the Cambridge Philosophical Society*, (21), 30–40.
  - Demerec, M., & Fano, U. (1945). Bacteriophage-Resistant Mutants in *Escherichia Coli*. *Genetics*, 30(2), 119–136.
  - Dodd, I. B., Perkins, A. J., Tsemitsidis, D., & Egan, J. B. (2001). Octamerization of lambda CI repressor is needed for effective repression of. *Genes & Development*, 15, 3013–3022.
  - Dunn, J. J., Studier, F. W., & Gottesman, M. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Journal of Molecular Biology*, 166(4), 477–535.
  - Edwards, A. L., & Batey, R. T. (2010). RiboSwitches: A Common RNA Regulatory Element. *Nature Education*, (3), 9. Retrieved from <https://www.nature.com/scitable/topicpage/riboswitches-a-common-rna-regulatory-element-14262702/>
  - Ellis, E. L. , Delbruck, M. (1940). the Growth of Bacteriophage and Lysis of the Host. *The Journal of General Physiology*, 23(5), 643–660.
  - Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step,

- precision cloning method with high throughput capability. *PLoS ONE*, 3(11), 1–7.
- Erdmann, V. A., Barciszewska, M. Z., Hochberg, A., de Groot, N., & Barciszewski, J. (2001). Regulatory RNAs. *Cell and Molecular Life Sciences*, 58(7), 960–977.
  - Esvelt, K. M., Carlson, J. C., & Liu, D. R. (2011). A System for the Continuous Directed Evolution of Biomolecules. *Nature*, 472(7344), 499–503.
  - Fehér, T., Karcagi, I., Blattner, F. R., & Pósfai, G. (2012). Bacteriophage recombineering in the lytic state using the lambda red recombinases. *Microbial Biotechnology*, 5(4), 466–476.
  - Fernandez-Gacio, A., Uguen, M., & Fastrez, J. (2003). Phage display as a tool for the directed evolution of enzymes. *Trends in Biotechnology*, 21(9), 408–414.
  - Findeiß, S., Etzel, M., Will, S., Mörl, M., & Stadler, P. F. (2017). Design of artificial riboswitches as biosensors. *Sensors (Switzerland)*, 17(9), 1–28.
  - Friedberg, E. C., Fischhaber, P. L., & Kisker, C. (2001). Error-Prone DNA Polymerases. *Cell*, 107(1), 9–12.
  - Galizi, R., & Jaramillo, A. (2019). Engineering CRISPR guide RNA riboswitches for in vivo applications. *Current Opinion in Biotechnology*, 55, 103–113.
  - Garcia, L. R., & Molineux, I. a N. J. (1995). Incomplete entry of bacteriophage T7 DNA into F plasmid-containing *Escherichia coli*. *Journal of Bacteriology*, 177(14), 4077–4083.
  - Garneau, J. R., Depardieu, F., Fortier, L. C., Bikard, D., & Monot, M. (2017). PhageTerm: A tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Scientific Reports*, 7(1), 1–10.
  - Gibbs, A. J. (2013). Viral taxonomy needs a spring clean; Its exploration era is over. *Virology Journal*, 10(1), 1–4.
  - Golkar, Z., Bagasra, O., & Gene Pace, D. (2014). Bacteriophage therapy: A potential solution for the antibiotic resistance crisis. *Journal of Infection in Developing Countries*, 8(2), 129–136.

- Gong, S., Wang, Y., Wang, Z., & Zhang, W. (2017). Computational methods for modeling aptamers and designing riboswitches. *International Journal of Molecular Sciences*, 18(11), 1–19.
- Gorzelnik, K. V., Cui, Z., Reed, C. A., Jakana, J., Young, R., & Zhang, J. (2016). Asymmetric cryo-EM structure of the canonical allovivivirus Q $\beta$  reveals a single maturation protein and the genomic ssRNA in situ. *Proceedings of the National Academy of Sciences of the United States of America*, 113(41), 11519–11524.
- Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C., & Gelbart, W. M. (2000). Transduction. In W. H. Freeman (Ed.), *An Introduction to Genetic Analysis* (7th ed.). New York. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK21760/%0A>
- Groher, F., & Suess, B. (2014). Synthetic riboswitches - A tool comes of age. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839(10), 964–973.
- Guzman, L. M., Belin, D., Carson, M. J., & Beckwith, J. (1995). Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose PBAD Promoter. *Journal of Bacteriology*, 177(14), 4121–4130.
- Haber, J. E. (2008). *Evolution of Models of Homologous Recombination* BT - *Recombination and Meiosis: Models, Means, and Evolution*. (R. Egel & L. Dirk-Henner, Eds.) (Vol. 3). Berlin, Heidelberg: Springer Berlin/Heidelberg.
- Hankin, E. H. (1896). L'action bactericide des eaux de la Jumna et du Gange sur le vibron du cholera. *Annales de l'Institut Pasteur (in French)*. 10: 511–523., (10), 511–523.
- Hendrix, R. W. (2003). Bacteriophage genomics. *Current Opinion in Microbiology*, 6(5), 506–511.
- Hermann, T., & Patel, D. J. (2000). Adaptive recognition by nucleic acid aptamers. *Science*, 287(5454), 820–825.
- Higgs, P. G. (2000). RNA secondary structure: physical and computational aspects. *Quarterly Reviews of Biophysics*, 33(3), 199–253.
- Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., & Charpentier, E.

- (2018). The Biology of CRISPR-Cas: Backward and Forward. *Cell*, 172(6), 1239–1259.
- Hoelz, D. J., Hickey, R. J., & Malkas, L. H. (2004). DNA Replication: Prokaryotes and Yeast. *Encyclopedia of Life Sciences*, 1–9.
  - Hoinka, J., Backofen, R., & Przytycka, T. M. (2018). AptaSUITE: A Full-Featured Bioinformatics Framework for the Comprehensive Analysis of Aptamers from HT-SELEX Experiments. *Molecular Therapy - Nucleic Acids*, 11(June), 515–517.
  - Huang, D., Holtz, W. J., & Maharbiz, M. M. (2012). A genetic bistable switch utilizing nonlinear protein degradation. *Journal of Biological Engineering*, 6, 1–13.
  - Huang, L., Yuan, Z., Liu, P., & Zhou, T. (2015). Effects of promoter leakage on dynamics of gene expression. *BMC Systems Biology*, 9(1), 1–12.
  - Hubbard, B. P., Badran, A. H., Zuris, J. A., Guilinger, J. P., Davis, K. M., Chen, L., ... Liu, D. R. (2015). Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nature Methods*, 12(10), 939–942.
  - Hyman, P., & Abedon, S. T. (2009). *Practical methods for determining phage growth parameters*. (M. R. J. Clokie & A. M. Kropinski, Eds.), *Bacteriophages, Methods and Protocols* (Vol. 1: Isolation). Humana Press.
  - Isalan, M. (2006). Construction of semi-randomized gene libraries with weighted oligonucleotide synthesis and PCR. *Nature Protocols*, 1(1), 468–475.
  - Jenison, R. D., Gill, S. C., Pardi, A., & Polisky, B. (1994). High-Resolution Molecular Discrimination by RNA. *Science*, 263, 1425–1429.
  - Jiang, F., & Doudna, J. A. (2017). CRISPR – Cas9 Structures and Mechanisms. *Annual Review of BioPhysics*, 46, 505–529.
  - Jiang, W., Maniv, I., Arain, F., Wang, Y., Levin, B. R., & Marraffini, L. A. (2013). Dealing with the Evolutionary Downside of CRISPR Immunity: Bacteria and Beneficial Plasmids. *PLoS Genetics*, 9(9), 1–13.
  - Kashiwagi, A., & Yomo, T. (2011). Ongoing phenotypic and genomic changes in experimental coevolution of rna bacteriophage q $\beta$  and escherichia coli. *PLoS Genetics*, 7(8), 1–13.



- Krishnamurthy, M., Hennelly, S. P., Dale, T., Starkenburg, S. R., Martí-Arbona, R., Fox, D. T., ... Unkefer, C. J. (2015). Tunable Riboregulator Switches for Post-transcriptional Control of Gene Expression. *ACS Synthetic Biology*, 4(12), 1326–1334.
- Kuchment, A. (2012). *The Forgotten Cure: The Past and Future of Phage Therapy*. New York: Springer US.
- Kutter, E., De Vos, D., Gvasalia, G., Alavidze, Z., Gogokhia, L., Kuhl, S., & Abedon, S. (2010). Phage Therapy in Clinical Practice: Treatment of Human Infections. *Current Pharmaceutical Biotechnology*, 11(1), 69–86.
- Lakhin, A. V., Tarantul, V. Z., & Gening, L. V. (2013). Aptamers: Problems, solutions and prospects. *Acta Naturae*, 5(19), 34–43.
- Lee, W. J., Billington, C., Hudson, J. A., & Heinemann, J. A. (2011). Isolation and characterization of phages infecting *Bacillus cereus*. *Letters in Applied Microbiology*, 52(5), 456–464.
- Liu, C. C., Qi, L., Lucks, J. B., Segall-shapiro, T. H., Wang, D., Mutalik, V. K., & Arkin, A. P. (2012). An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nature Methods*, 9(11), 1088–1094.
- Liu, H., & Yu, J. (2018). *Challenges of SELEX and Demerits of Aptamer-Based Methods*. In *Aptamers for Analytical Application*. (Y. Dong, Ed.), *Aptamers for Analytical Applications* (1st ed.). Wiley-VCH.
- Liu, J., Dehbi, M., Moeck, G., Arhin, F., Banda, P., Bergeron, D., ... DuBow, M. (2004). Antimicrobial drug discovery through bacteriophage genomics. *Nature Biotechnology*, 22(2), 185–191.
- Lynch, S. A., Desai, S. K., Sajja, H. K., & Gallivan, J. P. (2007). A High Throughput Screen for Synthetic Riboswitches Reveals Mechanistic Insights into their Function. *Chemical Biology*, 14(2), 173–184. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/12999698>
- Lynch, S. A., & Gallivan, J. P. (2009). A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Research*, 37(1), 184–192.
- Machtel, P., Bąkowska-Żywicka, K., & Żywicki, M. (2016). Emerging applications of riboswitches – from antibacterial targets to molecular tools. *Journal of Applied Genetics*, 57(4), 531–541.

- Mahr, R., Gätgens, C., Gätgens, J., Polen, T., Kalinowski, J., & Frunzke, J. (2015). Biosensor-driven adaptive laboratory evolution of l-valine production in *Corynebacterium glutamicum*. *Metabolic Engineering*, 32, 184–194.
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., ... Koonin, E. V. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, 9(6), 467–477.
- Malmstrom, C. (2010). Ecologists Study the Interactions of Organisms and Their Environment. *Nature Education Knowledge*, 3(10)88.
- Margoliash, E. (1963). Primary Structure and Evolution of Cytochrome C. *Proceedings of the National Academy of Sciences of the United States of America*, 50(1961), 672–679.
- Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR interference. *Science*, 322(5909), 1843–1845.
- Mehdizadeh, E., Saeid, M., & Barzegar, A. (2016). Riboswitches : From living biosensors to novel targets of antibiotics. *Gene*, 592(2), 244–259.
- Mehrotra, P. (2016). Biosensors and their applications – A review. *Journal of Oral Biology and Craniofacial Research*, 6(2), 153–159.
- Mi, Y., Liu, Y., Rabbani, Z. N., Yang, Z., Urban, J. H., Sullenger, B. A., & Clary, B. M. (2010). In vivo selection of tumor-targeting RNA motifs. *Nature Chemical Biology*, 6(1), 22–24.
- Miranda-Ríos, J., Navarro, M., & Sobero, M. (2001). A conserved RNA structure ( thi box ) is involved in regulation of thiamin biosynthetic gene. *Proceedings of the National Academy of Science of the United States of America*, 98(17), 9736–9741.
- Molineux, I. J. (2005). Chapter 20. The T7 group. (R. Lane, Ed.), *The Bacteriophages* (2nd ed.). Oxford: Oxford University Press.
- Montange, R. K., & Batey, R. T. (2006). Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, 441(7097), 1172–1175.
- Morris, K. V., & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews Genetics*, 15(6), 423–437.
- Moye, Z. D., Woolston, J., & Sulakvelidze, A. (2018). Bacteriophage



- applications for food production and processing. *Viruses*, 10(4), 1–22.
- Nakahira, Y., Ogawa, A., Asano, H., Oyama, T., & Tozawa, Y. (2013). Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in cyanobacterium *synechococcus elongatus* PCC 7942. *Plant and Cell Physiology*, 54(10), 1724–1735.
  - Naldini, L., Blömer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F. H., ... Trono, D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science*, 272(5259), 263–267.
  - Neidhart, F. C., Schaechter, M., & Ingraham, John L. (1990). *Physiology of the bacterial cell: a molecular approach*. (Sinauer Associates, Ed.). Sunderland, Mass.
  - Nudler, E., & Mironov, A. S. (2004). The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences*, 29(1), 11–17.
  - O'Rourke, S., Wietzorrek, A., Fowler, K., Corre, C., Challis, G. L., & Chater, K. F. (2009). Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in *Streptomyces coelicolor*. *Molecular Microbiology*, 71(3), 763–778.
  - Orlov, M. A., Ryasik, A. A., & Sorokin, A. A. (2018). Destabilization of the DNA Duplex of Actively Replicating Promoters of T7-Like Bacteriophages. *Molecular Biology*, 52(5), 686–692.
  - Pearce, T. (2010). From “circumstances” to “environment”: Herbert Spencer and the origins of the idea of organism-environment interaction. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 41(3), 241–252.
  - Pham, H. L., Wong, A., Chua, N., Teo, W. S., Yew, W. S., & Chang, M. W. (2017). Engineering a riboswitch-based genetic platform for the self-directed evolution of acid-tolerant phenotypes. *Nature Communications*, 8(1), 1–12.
  - Pires, D. P., Cleto, S., Sillankorva, S., Azeredo, J., & Lu, T. K. (2016). Genetically Engineered Phages: a Review of Advances over the Last. *Microbiology and Molecular Biology Reviews*, 80(3), 523–543.

- Qimron, U., Marintcheva, B., Tabor, S., & Richardson, C. C. (2006). Genomewide screens for *Escherichia coli* genes affecting growth of T7 bacteriophage. *Proceedings of the National Academy of Sciences of the United States of America*, 103(50), 19039–19044.
- Ramajani, S., & Sayre, R. T. (2018). Biosensors for the Detection and Quantification of AI-2 Class Quorum-Sensing Compounds. In L. Leoni & G. Rampioni (Eds.), *Quorum Sensing. Meth Mol Biol* vol 1673 (pp. 77–88). New York: Humana Press.
- Rankin, C. J., Fuller, E. N., Hamor, K. H., Gabarra, S. A., & Shields, T. P. (2006). A Simple Fluorescent Biosensor for Theophylline Based on its RNA Aptamer. *Nucleosides, Nucleotides and Nucleic Acids*, 25(12), 1407–1424.
- Ravikumar, S., Baylon, M. G., Park, S. J., & Choi, J. il. (2017). Engineered microbial biosensors based on bacterial two-component systems as synthetic biotechnology platforms in bioremediation and biorefinery. *Microbial Cell Factories*, 16(1), 1–10.
- Retwitzer, M. D., Kifer, I., Sengupta, S., & Yakhini, Z. (2015). An Efficient Minimum Free Energy Structure- Based Search Method for Riboswitch Identification Based on Inverse RNA Folding. *PLoS ONE*, 10(7), 1–18.
- Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suit. *Trends in Genetics*, 16(6), 276–277.
- Robinson, C. J., Medina-Stacey, D., Wu, M. C., Vincent, H. A., & Micklefield, J. (2016). Rewiring Riboswitches to Create New Genetic Circuits in Bacteria. *Methods in Enzymology*, 575(May), 319–348.
- Rodrigo, G., & Jaramillo, A. (2014). RiboMaker : computational design of conformation-based riboregulation. *Bioinformatics*, 30(17), 2508–2510.
- Rodrigo, G., Landrain, T. E., & Jaramillo, A. (2012). De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. *Proceedings of the National Academy of Sciences of the United States of America*, 109(38), 15271–15276.
- Rogers, J. K., Guzman, C. D., Taylor, N. D., Raman, S., Anderson, K., & Church, G. M. (2015). Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Research*, 43(15), 7648–

7660.

- Rossi, R., Montecucco, A., Ciarrocchi, G., & Biamonti, G. (1997). Functional characterization of the T4 DNA ligase: A new insight into the mechanism of action. *Nucleic Acids Research*, 25(11), 2106–2113.
- Roth, A., & Breaker, R. R. (2009). The Structural and Functional Diversity of Metabolite-binding Riboswitches. *Annual Review of Biochemistry*, (78), 305–334.
- Robinson, D. A., Dillon, C. P., Kwiatkowski, A. V., Sievers, C., Yang, L., Kopinja, J., ... Van Parijs, L. (2003). A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nature Genetics*, 33(3), 401–406.
- Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 27(10), 946–952.
- Schmitt, C. K., & Molineux, I. J. (1991). Expression of gene 1.2 and gene 10 of bacteriophage T7 is lethal to F plasmid-containing *Escherichia coli*. *Journal of Bacteriology*, 173(4), 1536–1543.
- Schofield, D., Sharp, N. J., & Westwater, C. (2012). Phage-based platforms for the clinical detection of human bacterial pathogens. *Bacteriophage*, 2(2), 105–121.
- Schweet, R., Lamfrom, H., & Allen, E. (1958). The Synthesis of Hemoglobin in a Cell-free System. *Proceedings of the National Academy of Sciences*, 44, 1029–1035.
- Serganov, A., & Nudler, E. (2013). A Decade of RiboSwitches. *Cell*, 152(1–2), 17–24.
- Shaw, W. V., & Brodsky, R. F. (1968). Characterization of chloramphenicol acetyltransferase from chloramphenicol-resistant *Staphylococcus aureus*. *Journal of Bacteriology*, 95(1), 28–36.
- Shen, S., Rodrigo, G., Prakash, S., Majer, E., Landrain, T. E., Jaramillo, A., & Kirov, B. (2015). Dynamic signal processing by ribozyme-mediated RNA circuits to control gene expression. *Nucleic Acids Research*, 43(10), 5158–5170.
- Sherwood, A. V., & Henkin, T. M. (2016). Riboswitch-Mediated Gene

Regulation : Novel RNA Architectures Dictate Gene Expression Responses. *Annual Review of Microbiology*, (70), 361–374.

- Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., & Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nature Biotechnology*, 19(8), 751–755.
- Sidda, J. D., Song, L., Poon, V., Al-Bassam, M., Lazos, O., Buttner, M. J., ... Corre, C. (2014). Discovery of a family of  $\gamma$ -aminobutyrate ureas via rational derepression of a silent bacterial gene cluster. *Chemical Science*, 5(1), 86–89.
- Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., De Las Heras, A., ... De Lorenzo, V. (2013). The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Research*, 41(D1), 666–675.
- Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228(4705), 1315–1317.
- Smith, G. P., & Petrenko, V. A. (1997). Phage display. *Chemical Reviews*, 97(2), 391–410.
- Snyder, L., Peter, J. E., Henkin, T. M., & Champness, W. (2013). Lysogeny: the  $\lambda$  Paradigm and the Role of Lysogenic Conversion in Bacterial Pathogenesis. In *Molecular Genetics of Bacteria* (4th ed., pp. 340–343). Washington, DC: ASM Press.
- Springman, R., Keller, T., Molineux, I. J., & Bull, J. J. (2010). Evolution at a high imposed mutation rate: Adaptation obscures the load in phage T7. *Genetics*, 184(1), 221–232.
- Stella, R. G., Wiechert, J., Noack, S., & Frunzke, J. (2019). Evolutionary engineering of *Corynebacterium glutamicum*. *Biotechnology Journal*, 14(9), 1800444 (1–10).
- Stoltenburg, R., Reinemann, C., & Strehlitz, B. (2007). SELEX – A ( r ) evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering*, 24, 381–403.
- Storms, Z. J., Teel, M. R., Mercurio, K., & Sauvageau, D. (2019). The

Virulence Index : A Metric for Quantitative Analysis of Phage Virulence. *Phage*, 1(1), 17–26.

- Studier, F. W., & Dunn, J. J. (1982). Organization and expression of bacteriophage T7 DNA. *Cold Spring Harbor Symposia on Quantitative Biology*, 47(2), 999–1007.
- Sudarsan, N., Barrick, J. E., & Breaker, R. R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA*, (203), 644–647.
- Suess, B., Fink, B., Berens, C., Stentz, R., & Hillen, W. (2004). A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Research*, 32(4), 1610–1614.
- Sugimura, T., Nagao, M., & Okada, Y. (1966). Carcinogenic Action of N-Methyl-N'-nitro- N-nitrosoguanidine. *Nature*, 210(5039), 962–963.
- Suzuki, T., Miller, C., Guo, L. T., Ho, J. M. L., Bryson, D. I., Wang, Y. S., ... Söll, D. (2017). Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nature Chemical Biology*, 13(12), 1261–1266.
- Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., & Voigt, C. A. (2012). Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Research*, 40(17), 8773–8781.
- Tinoco, I., & Bustamante, C. (1999). How RNA folds. *Journal of Molecular Biology*, 293(2), 271–281.
- Tolkien, J. R. R. (1954). The Lord of The Rings. In *The Fellowship of the Ring* (p. 87). London: George Allen & Unwin.
- Topp, S., & Gallivan, J. P. (2008). Guiding Bacteria with Small Molecules and RNA. *Journal of the American Chemical Society*, 129(21), 6807–6811.
- Tucker, B. J., & Breaker, R. R. (2005). Riboswitches as versatile gene control elements. *Current Opinion in Structural Biology*, 15(3 SPEC. ISS.), 342–348.
- Twort, F. W. (1915). An Investigation on the Nature of Ultra-Microscopic Viruses. *The Lancet*, 186(4814), 1241–1243.
- Vallée-Bélisle, A., & Plaxco, K. W. (2010). Structure-switching biosensors: Inspired by Nature. *Current Opinion in Structural Biology*, 20(4), 518–526.
- Vilar, J. M. G., & Saiz, L. (2010). CplexA: A Mathematica package to study

- macromolecular-assembly control of gene expression. *Bioinformatics*, 26(16), 2060–2061. <https://doi.org/10.1093/bioinformatics/btq328>
- Wagner, T. E., Becraft, J. R., Bodner, K., Teague, B., Zhang, X., Woo, A., ... Weiss, R. (2018). Small-molecule-based regulation of RNA-delivered circuits in mammalian cells. *Natural Chemical Biology*, 40(11), 1043–50.
  - Wang, B., Barahona, M., & Buck, M. (2013). A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals. *Biosensors and Bioelectronics*, 40(1), 368–376.
  - Wang, H., Zhang, Y., Yang, H., Qin, M., Ding, X., Liu, R., & Jiang, Y. (2018). In Vivo SELEX of an Inhibitory NSCLC-Specific RNA Aptamer from PEGylated RNA Library. *Molecular Therapy - Nucleic Acids*, 10(March), 187–198.
  - Weiss, R. (2001). *Cellular computation and communications using engineered genetic regulatory networks*. Retrieved from <http://dspace.mit.edu/handle/1721.1/8228>
  - Weyant, R. S., Edmonds, P., & Swaminathan, B. (1990). Effect of Ionic and Nonionic detergents in Taq polymerase. *Biotechniques*, 9(3), 308–309. Retrieved from <https://www.thermofisher.com/order/catalog/product/K1081>
  - Win, M. N., & Smolke, C. D. (2008). Higher-Order Cellular Information Processing with Synthetic RNA Devices. *Science*, 322(5900), 456–460.
  - Wrenbeck, E. E., Klesmith, J. R., Sapleton, J. A., Adeniran, A., Tyo, K. E., & Whitehead, T. A. (2016). Plasmid-based one-pot saturation mutagenesis. *Nature Methods*, (11), 928–930.
  - Xie, Y., Wahab, L., & Gill, J. J. (2018). Development and validation of a microtiter plate-based assay for determination of bacteriophage host range and virulence. *Viruses*, 10(4), 1–16. <https://doi.org/10.3390/v10040189>
  - Yosef, I., Goren, M. G., Globus, R., Molshanski-Mor, S., & Qimron, U. (2017a). (Suppl)Extending the Host Range of Bacteriophage Particles for DNA Transduction. *Molecular Cell*, 66(5), 721–728.
  - Yosef, I., Goren, M. G., Globus, R., Molshanski-Mor, S., & Qimron, U.

- (2017b). Extending the Host Range of Bacteriophage Particles for DNA Transduction. *Molecular Cell*, 66(5), 721–728.
- You, L., Tong, R., Li, M., Liu, Y., Xue, J., & Lu, Y. (2019). Advancements and Obstacles of CRISPR-Cas9 Technology in Translational Research. *Molecular Therapy - Methods and Clinical Development*, 13(June), 359–370.
  - Zinder, N. D., & Lederberg, J. (1952). Genetic exchange in Salmonella. *Journal of Bacteriology*, 64(5), 679–699. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12999698><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC169409>